DOCUMENT RESUME

ED 297 442 EA 020 170

AUTHOR Foote, Tom H.
TITLE The Managerial Grid for Teachers: Evidence, Practical Applications, and Directions for Future Research.
PUB DATE 28 Feb 88
NOTE 63p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Dropout Rate; Elementary Secondary Education; *Relevance (Education); *School Effectiveness; *Student Evaluation of Teacher Performance; *Student Needs; Teacher Education; *Teacher Effectiveness

ABSTRACT

Increasing dropout rates, marked decline in college preparatory enrollment, and reports of students' growing dissatisfaction with school prompted this exploration of the dimensions underlying teaching and the ways that microeconomics might positively influence teaching effectiveness. The literature review supports 10 advantages offered by student evaluations of teaching (SETS) in complementing student achievement as a schooling effectiveness measure. The Fox SETS studies show that teachers' enthusiasm accounts for twice the student achievement that lecture content does. This study's analysis was based on a sample of 60 teacher aggregates of 3,961 student questionnaires gathered over 2 years from the same high school. Factor analysis of 52 variables revealed "student-caring" versus "task-driven" as key research-based, teaching dimensions. The resulting "managerial grid" for teachers offered graphic perspectives of interrelationships not afforded by conventional frameworks. Using "student satisfaction with teacher performance" as the dependent variable, allocative efficiency was probed using planar and other quantitative techniques. Results show that "student-caring" has substantially more influence on satisfaction than does "task-driven." Teachers are implicitly paid considerably less for caring than for task-driven behaviors. Reallocating salaries to produce teaching that interests students poses formidable technical and political problems. However, more caring for students, not necessarily more money allocated to it, should also produce higher percentages of satisfied students. The managerial grid should help define and achieve more appropriate teaching styles. Included are 128 endnotes. (Author/MLH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED297442

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Tom H. Foote

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

THE MANAGERIAL GRID FOR TEACHERS: EVIDENCE, PRACTICAL APPLICATIONS, AND DIRECTIONS FOR FUTURE RESEARCH

Tom H. Foote, Assistant Professor

Department of Educational Administration

Teachers College, Columbia University

February 28, 1988

Division A Paper Number 101
1988 AERA Annual Meeting in New Orleans
To be presented as part of Session 49.18:
"Career Aspirations and Career Structures"
Saturday 8:15-10:15 A.M.
Sheraton Grand Ballroom B, 5th Floor

Please direct comments to:

Tom H. Foote
Box 13
Teachers College, Columbia University
New York, NY 10027
Telephones:
Office 212/678-3857 or 3726
Home/Recorder 212/678-1555

EA 020 170

ABSTRACT

PROBLEM/OBJECTIVES

Precipitous increases in high school dropout rates, marked declines in college preparatory enrollment, and reports that students are dissatisfied with school, prompted this exploration of what dimensions underlie teaching and how microeconomics might help efficiently produce teaching that genuinely interests, and challenges students.

PERSPECTIVES/THEORETICAL FRAMEWORK

The literature review finds support for ten advantages that student evaluations of teaching (SETs) offer in complementing student achievement as a schooling effectiveness measure. And reanalysis here of the Dr. Fox studies of SETs shows teacher's enthusiasm to not only strongly affect evaluations, but also to consistently account for over twice the student achievement that lecture content does.

METHODS/DATA SOURCE

This study's main analysis, however, was based on a sample of 60 teacher aggregates of 3,961 student questionnaires gathered over two years from the same high school. Factor analysis of 52 variables revealed "student-caring" versus "task-driven," as key teaching dimensions--dimensions that persist across fourteen cited management and psychology studies of other supervisory contexts. The resulting "managerial grid for teachers" offered graphic perspectives of interrelationships not afforded by conventional frameworks.

Using a proxy for "student satisfaction with teacher performance" as the dependent variable, allocative efficiency was probed with planar, quadratic, Cobb-Douglas, probit and logit models of educational production functions fitted to proxies for the two grid dimensions. Though the curved surface of the quadratic explained the most variation and reflected textbook-like diminishing returns along both dimensions, the planar model best satisfied constraints of fit, and simplicity. Adding teacher's salary determinants and controlling statistically for nonteacher influences produced substantially unequal numerators and denominators for the requisite marginal-product-to-marginal-price ratios.

RESULTS/CONCLUSIONS

According to microeconomic theory, these disparities imply that the high school observed does not efficiently allocate salaries to produce teaching that interests students. Student-caring has substantially more influence on satisfaction than does task-driven but teachers are <u>implicitly</u> paid considerably <u>less</u> for caring than for task-driven behaviors. Moreover, teacher's salary is <u>explicitly</u> driven most by teaching experience, and relatively less by education--neither of which relate to student perceptions of teacher performance (just as others, e.g., Hanushek, have found them not to relate to student achievement).

IMPORTANCE/POLICY IMPLICATIONS

Reallocating salaries to efficiently produce teaching that interests students surely poses formidable technical and political problems. Yet the data support the intuitively unsurprising notion that more student-caring--not necessarily more money reallocated to it--should also produce higher percentages of satisfied students. To that end, changes are suggested involving teacher-training institutions, hiring, career ladders, in-service enrichment programs, and teacher pay.

Policy change responsive to this study is well within reach in California. For example, just as the California Assessment Program (CAP) gathers achievement measures statewide, so too--perhaps even on the same test instrument--could it gather SET data covering teacher traits and student satisfaction. Indeed, just as California monetarily rewards CAP achievement gains, so too might it reward improvements in the satisfaction and retention of students.

SCIENTIFIC IMPORTANCE/FUTURE RESEARCH

Future research is urged that uses the managerial grid as the floor of a three dimensional space to: 1) analyze more comprehensive models involving interaction of achievement, satisfaction, teaching styles and other key schooling variables; 2) establish quantitatively, the output consequences of teaching style trade-offs; and, 3) articulate preferred teaching styles (as Blake et al. did in applying their grid to other professions).

Above all, the grid should help define and achieve teaching styles that better mix student-caring/task-driven inputs, and satisfaction/achievement outputs--to aid the bigger picture of long-term student wellbeing and reversal of trends that prompted this study.

[Note: This work builds upon: earlier research done at Stanford University, funded by the Ford Foundation. Although I alone, am responsible for any errors or omissions, I remain especially grateful to Ed Bridges, Jay Chambers, Nathan L. Gage, Mike Garet, Michael Kirst, Henry Levin, Ray Perry, and others for their good ideas and feedback.]

1.1 PROBLEM

In April, 1983, the National Commission on Excellence in Education (NCEE) reported that the percentage of students taking "general track" (i.e., non-college preparatory and non-vocational) courses jumped from 12 percent in 1964 to 42 percent in 1979.[1] In California, the percentage of students who drop out between the ninth and twelfth grades soared from 12 percent in 1970, to 22 percent in 1979.[2] This is consistent with an earlier report of the California Commission for Reform of Intermediate and Secondary Education (RISE Commission, 1975) that stated "Increasing numbers of young people find schooling boring and ineffective, unable to challenge their abilities, meet their goals, or prepare them for adulthood."[3] It also agrees with a 1983 Carnegie Foundation study that states "Most seniors agree that there should be more emphasis on academic subjects and that poor teaching has interfered with their education . . . "[4] Indeed, the rash of 1983 reports seem to agree that "teacher training should be revised; that there should be more emphasis on subject matter . . . and . . . better approaches to teaching methods."[5]

More recent reports have reaffirmed the above. The Carnegie Forum on Education and the Economy, states:

> Concern over the quality of education in this country has been expressed in repeated warnings from the Education Commission of the States' Task force on Education and economic Growth, the President's Commission on Industrial Competitiveness, the National Alliance of Business and others...Large numbers of American children are in limbo--ignorant of the past and unprepared for the future. Many are dropping out--not just out of school but out of productive society...An economy based on people who think for a living requires schools dedicated to the creation of environments in which students become very adept at think for themselves, places where they master the art of learning and acquire a strong taste for it.[6]

If the above trends are as severe as reported, they may be due, in part, to the lack of any <u>systematic</u> plan to produce teaching that attracts, challenges, and fulfills students. More precisely, just as changing teachers' salary determinants could produce more student learning per dollar,[7] so too might a similar change yield more efficient production of teaching that engages and enlivens students.[8]

The RISE, NCEE, National Task Force on Education for Economic Growth (NTFEEG), and Carnegie reports, recommend changes consistent with such a plan; e.g., they urge that teacher's salaries[9] and jobs[10] be linked to teacher performance; that learners be considered "the primary clients--the most important individuals--served by the schools";[11] and, that staff evaluation procedures ". . . allow for the consideration of representative views of those affected by the performance of the staff member under review."[12] The Carnegie report adds that present incentives:

> ...not only do not reward performance and productivity, but sometimes actually discourage them. Americans already spend more per capita on education than any other country in the free world. This plan will work only if those who must finance it are convinced that the funds are being used as efficiently as possible.[13]

How to use those funds as efficiently as possible? If a school district chose to follow the above recommendations in a strict economic sense, that is, if--subject to a given budget constraint--it wanted to produce a maximum of teaching that truly engages and enriches students--then it would first need to ascertain the requisite microeconomic conditions of efficiency. Defining "efficiency" is aided by Levin's distinctions between technical, allocative, and social efficiency, where:

1. Technical efficiency refers to transforming a given combination of inputs into the maximum output (graphically depicted in two-input space by any combination of the two inputs on a given isoproduct curve);

2. Allocative or "price" efficiency considers relative input prices to find what technically efficient combination of inputs produces the maximum output for a given budget (Graphically, this is where the budget line is tangent to the highest possible isoquant)[14]; and, finally,

3. Social efficiency means lifting society to its highest possible indifference curve by "the optimal" combination of outputs (that point in utility space where the highest possible social welfare function is tangent to the grand utility possibility frontier).[15]

While this study speaks directly to allocative or price efficiency (in allocating teachers' salaries to produce maximum student satisfaction), it also relates to views on social efficiency. More specifically, some argue that society would be better off caring less about the traditional products (e.g., achievement) of schools and more about how the schooling process affects students and teachers. In his "Study of Schooling," John Goodlad concludes:

> We can make our schools more efficient. But making them relevant in the lives of boys and girls is one of the most demanding challenges we face.[16]

In a similar vein, Jencks et al. argue:

> Instead of evaluating schools in terms of long-term effects on their alumni, which appear to be relatively uniform, we think it wiser to evaluate schools in terms

> of their immediate effects on teachers and students, which appear much more variable. Some schools are dull, depressing, even terrifying places, while others are lively, comfortable, and reassuring. If we think of school life as an end in itself rather than a means to some other end, such differences would not do much to make adults more equal, but it would do a great deal to make the quality of children's (and teachers') lives more equal. Since children are in school for a fifth of their lives, this would be a significant accomplishment.[17]

Some prefer this focus even if it results in a classical product transformation curve--where gains in student satisfaction are traded for declines in student achievement. Indeed, A.S. Neil writes that he would rather see schools produce happy street cleaners than neurotic scholars.[18] Though related to Neil's view, the bulk of this study will be considerably less philosophical, focusing largely on the technical and empirical issues surrounding the allocative efficiency of producing schooling that interests students.[19]

1.2 QUESTIONS

The challenge of allocating all resources to efficiently produce satisfied students is simplified by seeing that teacher's salaries and student satisfaction with teaching cover most of the bigger picture linking overall district budgets to student satisfaction with school in general. This follows because teachers' salaries consume approximately 70 percent of those budgets; and, roughly 80 percent of the student's time in school is spent with teachers.

Measuring satisfaction becomes empirically feasible by ascertaining student perceptions of, and reactions to, teacher behaviors via "Student Evaluation of Teacher" (SET) questionnaires.

From the above perspective, this study attempts to answer the following:

1. What advantages might student evaluations of teachers offer California public high schools as an added measure of schooling effectiveness?

2. What fundamental dimensions of teaching underlie such evaluations?

3. To what extent does one public high school in California already satisfy the microeconomic conditions to efficiently produce teaching that interests students? That is, how much impact do specific teacher traits (e.g., knowledge of the subject matter, preparation for class, enthusiasm) have on overall student satisfaction? What implicit monetary prices do districts pay for those traits? To what extent are the impact-to-price ratios disparate across those traits?

4. To what specific inefficiencies do any disparities point and what policy changes do they imply?

2.1 PROBLEM CONTEXT[20]

Efficiently translating dollars into desired educational outcomes has long been a concern of economists. Writing over two centuries ago in his Wealth of Nations, Adam Smith urged that the school master be "partly, but not wholly paid by the public; because if he was wholly, or even principally paid by it, he would soon learn to neglect his business."[21] Such pessimism about the effects of unsound incentives in schools has been persistent. Indeed, economists' claims of misguided incentives inducing inefficiencies in today's schools, are widespread.[22] [23] [24] [25] [26] Most recently, Hanushek asserts: "The incentives existing within schools appear to provide a compelling explanation of public school inefficiencies"[27] and ". . . the current structure of public education offers little hope for self correction."[28] Levin articulates the broader problems:

> The conditions under which educational expenditures will translate into improved educational outcomes are fairly stringent. First, there must be substantial agreement on which outcomes are important; second, there must be knowledge of how added resources can be used to improve those outcomes; and third, those who are responsible for the educational process must have incentives to maximize the socially desirable outcomes. In each case, the actual situation seems to contrast with the assumptions that are necessary for added dollars to improve educational results . . . There is no agreement on educational priorities, no body of knowledge that can predict the effects of different school policies on educational outcomes, and no relation between staff incentives and the rhetorical goals of schools. This all suggests that school budgets are not likely to be translated efficiently into improved educational outcomes. Rather, the actual application of the increased support is likely to be determined by the power and interests of the decision makers themselves . . .[29]

Important among the interests of decision makers--student achievement aside--is maintaining smooth political sailing among the adults with whom they work.[30] [31]

Hanushek feels that evaluation difficulties further influence these interests at the expense of both student and community interests:

> . . . The fact that school policies are so hard to evaluate makes it easier for teachers and school officials to be influenced by their personal interests . . . Without direct evaluations, teachers and school officials need not confront the possibility that their privately preferred policies may be useless for the student and bad for the community.[32]

Though school administrators' reluctance to evaluate teachers results, in part, from technical difficulties in measuring schooling outcomes, it also follows from the basic structural characteristics of the public school setting; i.e., service to captive clientele in a non-competitive, non-profit environment that yields no rewards to decision makers for efficiency in production.[33] In discussing the corresponding administrator strategies, Chambers argues:

> Since they are unable to appropriate any pecuniary benefits to themselves and since they generally operate subject only to a minimal survival mechanism, school decision-makers find that the cost of inefficient behavior is relatively low. Their efforts, however, are directed toward the reduction of the psychic costs involved in the management of district operations. Based upon these motivations, there is little incentive for school administrators to obtain precise evaluations of performance of the school system. Any such evaluation of the outputs of the system could only serve to reveal possible deficiencies in their own managerial abilities and, thus, threaten their survival. The more abstruse the method of evaluation, the greater reliance the school board and the community must place in the judgment of these professional educators.[34]

According to Hanushek, administrators are likely to go beyond mere abstrusive evaluation to help teachers because both groups' interests often coincide:

> Administrators, who are typically drawn from the ranks of teachers, share many conceptions about what is "right." Moreover, it generally serves the interest of administrators to accept teachers' arguments about reducing class sizes, raising salaries, and increasing expenditures; such policies increase the administrators' domain, lessen conflicts with their employees, and ultimately must affect their own salaries.[35]

These and other coincidental interests,[36] and the absence of rewards to educators to improve schooling, heighten the chances for interests of the deciders to encroach on those of children.[37] Moreover, any excellence that emerges is likely to flow more despite such a system than because of it; i.e., likely to flow more from personal goals and drive <u>within</u> especially motivated <u>individuals</u>, than from <u>organizational</u> incentives <u>external</u> to them.[38]

2.2 ALLOCATION OF TEACHERS' SALARIES

The widespread time-honored system of paying teachers according to years and units, rather than performance, clearly favors administrative convenience and political tranquillity over incentives for efficiency. Years and units are administratively easy to measure and translate into higher salaries. The process

requires no consideration of substantial inter-job differences in supply and demand, workloads, or required competencies (despite the high costs of ignoring these differences). Above all, it requires no periodic, conflict-producing evaluations of teachers. Indeed, the process is "automatic" in the sense that each year, "action" is taken on salaries--they increase regardless of classroom performance--with little effort or political sacrifice from administrators.[39]

Although the above salary practice is nearly universal throughout our public schools,[40] a consistent body of research shows that it does not efficiently produce student achievement.[41] [42] [43] [44] Indeed, Hanushek found no relationship whatsoever between student achievement and teacher's years of service or post B.A. units:

> . . . the present set of hiring practices leads to an inefficient allocation of resources. The analysis indicates that teaching experience and graduate education do not contribute to gains in student achievement scores. Moreover, the characteristics that do matter are not highly correlated with these factors. Yet these attributes are being purchased by the school district. Since turnover is costly, some average experience level over one year would be reasonable. However, the current average of over eleven years is certainly excessive.[45]

After reviewing a far wider variety of inputs among 130 separate studies, Hanushek adds: "Within the range of current school operations, variations in expenditures or in any other commonly identified determinants of school "quality" bear no systematic relationship to variations in the performance of students."[46]

Though certain studies show teacher experience to nonetheless correlate positively with achievement,[47] [48] [49] the research of Keeler and McCall suggests that rather than high experience causing high achievement, "the major reason for the relation between experience and achievement is the ability of experienced teachers to get into the schools of the high achievers."[50] Because most districts' teachers are concentrated near the high ends of the years and units scales, the cost of these inputs represents a significant and periodically increasing national[51] expense.

The portion of this expense that serves to retain teachers can be justified because it reduces the costs of recruiting and training. And relative to a more competitive performance-based pay system, where older teachers might refrain from helping younger ones, the portion of years and units costs that encourage teamwork between old and young can also be justified. Yet whatever the magnitude of these, and perhaps other, justifiable portions, it is likely that their total falls short of the entirety spent on these inputs nationally; especially with years and units often accounting for more than half of a district's maximum teacher salary,[52] and teacher's salaries consuming between 60 and 80 percent of districts' total operating budgets.

2.3 BETTER ALLOCATION CRITERIA?

If we reject the current pay criteria because the evidence suggests it is inefficient, what better criteria are there? Hanushek's review of 130 input/output studies, mentioned earlier, reveals studies that do, and others that do not, show statistically significant effects of teacher inputs on student performance. Though these results suggest that no input bears any nationally generalizable relationship to student performance, this does not preclude that any one study might have accurately measured effects that work well for the particular students sampled. For example, Levin found, via linear[53] regression coefficients, that reallocating teachers' salaries according to teachers' verbal scores could yield up to ten times the student verbal achievement per dollar that present allocation to teachers' experience does.[54] And a later work by Levin showed teacher's verbal score to be four times as potent as teacher experience.[55] Though the Levin studies suggest that teacher's verbal score might be a superior input, neither study is sufficiently definitive to warrant widespread policy change.

While others argue that competition would force reallocation to more efficient inputs,[56 57 58] this study tackles schools' incentives, evaluation, and salary allocation problems within the existing non-competitive setting; by incorporating consumer (student) preferences into a microeconomic analysis of those problems. More precisely, just as paying teachers according to years and units appears inconsistent with efficiently producing student achievement, so too might it be inconsistent with efficiently producing teaching that attracts students. The lack of any agreement on even the sign of the correlation between teacher's years experience and student ratings of teaching quality (e.g., across the many studies reviewed by Costin et al.[59]) suggests this might be so. And just as Levin found an input that seems to produce learning more efficiently among a given student sample, so too might a similar methodology reveal inputs that produce satisfaction more efficiently.

Such a methodology would require measuring students' perceptions of, and reactions to, specific teacher traits. This is commonly done with student evaluations of teaching questionnaires (SETs). Such an effort assumes that SET results would offer valuable information not afforded by conventional achievement measures--an assumption explored in the following section.

2.4 SETS AS A COMPLEMENTARY MEASURE OF SCHOOLING OUTPUT

The use of SETs as one measure of teacher output is widespread at the college level[60]--and scant among high schools. This may result, in part, from there being more serious focus on promotion and tenure for college instructors, and therefore greater need to evaluate their teaching systematically. Since research on SETs in high schools is also scant, some of the following arguments necessarily draw from college-level studies.

Though most would agree that student <u>learning</u> should be the principal output of schools, research on student evaluations of <u>teaching</u> (SETs) shows that SETs complement student achievement

and other measures of school outputs with the following advantages:

1. SETs can measure dimensions of the schooling process, along with other dimensions of its products: Standardized tests "were not designed to reveal the variety of ways in which teaching and learning can be creative, favorably opportunistic, and uniquely meaningful to students."[61] SETs, on the other hand, can be designed to measure at least some of these many teaching process variables.

2. SETs reflect a unique and central viewpoint: Students see a broad range of their teacher's behaviors over months of exposure and personally know the students with whom they see the teacher interact. Furthermore, as the intended receivers of teaching, students are the only true first-hand sources regarding the receiving process and its effects.

3. SETs can reflect the quality of student life in school, which some believe is important in itself: Jencks et al., were cited above making this argument persuasively, i.e., urging that schools be evaluated for immediate effects on teachers and students rather than for long-term effects on alumni (see Section 1).[62] Surely with so much of our childhood spent in school, it makes sense to remove any unnecessary alienation or sadness that schools might produce. But schools are hardly likely to learn the existence, much less the source, of these ill feelings if students have no mechanism to reveal them.

4. SETs can measure noncognitive traits (e.g., enthusiasm) exhibited by teachers, and students. Recent research suggests that such traits are at least as important to students' later success as are cognitive ones: In their latest study, Who Gets Ahead, Jencks et al. find:

> Taken together, noncognitive measures explained at least as much of the variance in men's status and earnings as test scores did. While we could not isolate any single personality characteristic that was critical to success, we can say that the relevant traits are largely independent of both cognitive skills and parental status.[63]

With noncognitive traits so important to later success, it makes sense to begin measuring them and exploring their interrelationships. Indeed, SETs carefully designed to reflect noncognitive traits of both teachers and students, could aid research into the synergy between them.

5. The act of seeking feedback through SETs can be a tangible sign to students that teachers and officials seek and value students' views. Moreover, the extent to which those views are acted on, make this a step, albeit a small one, toward some democratization of school decision making.

6. SETs are highly reliable. And since they are less correlated with student characteristics (e.g., SES) than are achievement tests, they are fairer to teachers. In summarizing the empirical evidence on SETs, Gage writes:

> As for reliability, the results are almost uniformly happy. Averaging the ratings of about 20 students on a single item concerning a set of teachers, such as a rating of the clarity of the teacher's explanations, yields a mean rating that has a reliability of about .8 or .9. That is, these mean ratings would correlate about .8 or .9 with the mean ratings of another set of 20 equivalent students. This finding has been obtained with almost perfect consistency during the 50 years or so since Remmers (1929) first established. In short, the pooled ratings of 20 or more students reveal substantial agreement within a class as compared with the amount of variation between classes.[64]

This is consistent with "split class" reliabilities averaging .85 for SET factor scores found in junior-senior high-school classes by Veldman and Peck.[65]

Achievement tests, on the other hand, are ordinarily highly correlated with student characteristics. Glass writes:

> Aside from the irrelevance of much of the content of standardized achievement tests, their use in evaluating teachers is unjust. Nonrandomly constituted classes give teachers of brighter pupils an unfair advantage. This remains true whether the statistician calculates simple gains, residual gains, true gains, true residual gains, covariance adjustments, etc. . . . available evidence indicates that teachers' effects on pupils' knowledge are not reliably measured by such tests.[66]

Thus SETs are fairer to the teacher in the sense that they are less sensitive to between-student differences (or in the words of Gage, they "reveal substantial agreement within a class as compared with the amount of variation between classes"). That is, they are fairer than are achievement tests, because they are less sensitive to student differences (like student's socioeconomic status) clearly beyond the teacher's control.

7. SET responses correlate positively with evaluations done by teacher-colleagues and supervisors. At the University of Washington, Guthrie [67,68] found correlations ranging from 0.30 to 0.63 between students' ratings and teacher-colleague ratings of the same teachers. And at the junior-senior high school level, Veldman and Peck[69] show agreement between SET results and evaluations by teacher's supervisors. This is consistent with the following argument: Though teachers may question the validity of student ratings, teachers usually know which of their colleagues they'd prefer their own children to study under, and such preferences tend to agree with the very ratings the teachers discount.

8. SETs offer wide applicability across courses: Unlike standardized tests, one SET form can apply to a wide variety of school subjects, thus permitting control of instrument type across courses.

9. SETs may help improve instruction: Most instructors would acknowledge that certain student feedback can help improve instruction (especially those who, because of feedback, no longer block the view of their own chalkboard work).[70] And many studies[71 72 73 74 75] do show greater gains in student ratings for teachers who have received and acted on SET feedback (though rated improvement has yet to be proven equivalent to actual improvement).

10. SET's are valid in that they relate positively--though not strongly--to student achievement: The central challenge to the advocate of SET's as a measure of teaching effectiveness, is to disprove the likelihood of either of the two following errors:

a. The demanding teacher produces much achievement but receives low student evaluations (Type I error: good teacher rated as poor teacher);

b. The well-liked easy teacher produces little student achievement but gets high ratings (Type II error: poor teacher rated as excellent teacher).[76]

The most celebrated studies to this end are the series of "Dr. Fox studies."[77] These began when Naftulin, Ware, and Donnelly attempted to prove that a charismatic though nonsubstantive lecturer could receive high ratings even from highly educated professionals. Remarkably, a large number of the listeners--including psychiatrists, psychologists, and educators--did evaluate the Dr. Fox lecture favorably; even though, according to the authors, the lecture contained no substantive information. The authors conclude that this "suggests to the educator that the extent to which his students are satisfied with his teaching, and even the degree to which they feel they have learned, reflects little more than their illusions of having learned."[78] Yet without any effort whatsoever to test whether, or how much, the listeners may have learned, the authors conclude much more than their results support.

The authors' approach improved considerably in three later studies.[79 80 81] In short, those studies actually measured how student satisfaction and student achievement varied across different levels of teacher enthusiasm, and measurable lecture content. In the last of the above studies, they conclude ". . . the results observed to date suggest that student ratings of highly expressive instructors may not reflect two important dimensions of teaching effectiveness, namely, substantiveness of instruction and degree of student achievement."[82]

FIGURE 2.1
DR. FOX STUDIES: PROPORTION OF ACHIEVEMENT AND SATISFACTION STATISTICALLY ASSOCIATED WITH LECTURE CONTENT VERSUS TEACHER ENTHUSIASM AS MEASURED BY OMEGA-SQUARED[83]

PROPORTION ACHIEVEMENT CAPTURED BY:	STUDY 2 1975	STUDY 4[84] 1977
TEACHER'S ENTHUSIASM	0.09	0.13
LECTURE CONTENT	0.04	0.06
ENTHUSIASM/CONTENT RATIO	2.25	2.17
PROPORTION SATISFACTION CAPTURED BY:		
TEACHER'S ENTHUSIASM	0.24	0.24
LECTURE CONTENT	0.04	0.04
ENTHUSIASM/CONTENT RATIO	6.00	6.00

What they fail to report about their own data, however, is the remarkably consistent across-study pattern concerning the percentage of achievement and satisfaction attributable to lecture content versus teacher enthusiasm (as measured by omega-squared, the r-squared analog for ANOVA). Across both studies, variation in enthusiasm accounts for nearly twice (2.25 for Study 2 versus 2.17 for Study 4) the variation in achievement that variation in lecture content does. And again, across both studies, enthusiasm accounts for exactly 6.00 times the variation in satisfaction that content does. This surprising consistency suggests the intuitively appealing hypothesis that if lecture content is held constant and high, then how much of that content is learned, will depend strongly on how enthusiastically the teacher delivers it. This makes sense, since little is likely to be learned, regardless of how content-rich the lecture, if the lecturer fails to obtain the students' attention.

Though none of the Dr. Fox material or any other research categorically proves or disproves the likelihood of either the above Type I or II errors (a fact that should caution any interpretation of SET results), the most carefully conducted studies of the correlation between SET's and student achievement[85 86 87 88] do show, on average, low positive correlation between ratings of teachers' overall effectiveness and students' final exam score. In fact, these correlations (e.g., those of Sullivan and Skanes)[89] are higher than those found in meta-analyses seeking what teacher behaviors affect achievement.[90 91]

In meta-analysis of links between ratings and achievement, Cohen used 41 independent validity studies reporting on 68 separate multisection college courses. He found achievement

correlated an average of .43 with overall instructor rating, and 0.47 with the overall course rating. He concludes:

> We can be quite confident that the relationship between ratings and achievement described in this meta-analysis is characterized by what Bracht and Glass (1968) term "external validity." That is, the present findings can be generalized to different students, instructors, institutions, and subject matter areas . . . Based on the findings of the meta-analysis, we can safely say that student ratings of instruction are a valid index of instructional effectiveness. Students do a pretty good job of distinguishing among teachers on the basis of how much they have learned. Thus, the present study lends support to the use of ratings as one component in the evaluation of teaching effectiveness.[92]

Indeed, the majority who have seriously studied SETs (Cohen, Gage, Costin et al.) conclude that they do relate, though not strongly, to student achievement.

Do the teachers who are easy graders get the highest ratings? Some studies say that they do,[93] [94] others say that they do not,[95] [96] and still others that report mixed results.[97] Costin et al. suggest that those positive relationships that were found between rating and grade received, are likely to result from higher interest in the course triggering both higher ratings and higher grades[98] (i.e., not from desperate teachers buying higher ratings by giving higher grades).

2.5 SET'S: SUMMARY VIEWS

It must again be stressed that although this study focuses on high schools, the relative scarcity of high school SET research meant heavy reliance here on college-level studies.

In summarizing their extensive review of SET research (which, like this review, covered some high school SET studies, though mostly college-level research), Costin et al. made this overall assessment:

> A review of empirical studies indicates that students' ratings can provide reliable and valid information on the quality of courses and instruction. Such information can be of use to academic departments in constructing normative data for the evaluation of teaching and may aid the individual instructor in improving his teaching effectiveness.[99]

However, with specific regard to using SET's for personnel decisions, McKeachie cautions:

> Student ratings of teaching are related to teacher effectiveness as measured by the achievement of the teacher's students. Nevertheless this does not mean that student ratings are sufficient evidence of teaching effectiveness. Ideally one would gather evidence from a number of sources . . . Moreover, ratings should be

> obtained over several courses and several semesters before being admitted as data in the faculty evaluation process. A rating of a particular course in a particular semester may be influenced by special circumstances that affect its validity. Even when ratings from several courses are available, they need to be interpreted by peers or administrators who know something about the nature of the courses and students involved.[100]

Exactly which teacher traits affect students' overall satisfaction with the teacher? French-Lazovik found that at the college level, over 90 percent of the variation in student ratings of overall teacher effectiveness could be explained by student ratings on the following three teacher trait items: 1) Interprets abstract ideas and theories clearly; 2) Gets students interested in the subject matter; and 3) Has increased student's skills in thinking. Moreover, the relative importance of these and other items in predicting overall student perceived teaching effectiveness maintained ". . . surprising consistency across a 15-year time span on two quite different campuses with different student and faculty populations."[101]

At the junior-senior high level, Veldman and Peck's research showed a 38-item questionnaire to embody these five factors: 1) friendly, cheerful, admired; 2) knowledgeable, poised; 3) interesting, preferred; 4) strict control; and 5) democratic procedure.[102] Persisting over three separate semester's data sets (554 teachers total), this structure, like that found by French-Lazovik, appears steadfast with respect to time. (These factors also closely resemble those that emerged in this study's Sections 6, 7, and 8; namely "student-caring"; and "task-driven.")

In short, the French-Lazovik results, the many studies reviewed by Costin et al.,[103] and Cohen,[104], and Veldman and Peck's junior-senior high-school level research suggest that student criteria for teaching effectiveness are substantive and fairly stable over time.

3.1 METHODS, OVERVIEW

The main empirical question posed here involves ascertaining the extent one California public high school efficiently produces teaching that interests students. That question will be answered by gathering two years of SET data; adding to it teacher salaries and variables that determine salary; and using regression models to find the extent the data satisfy microeconomics' specific conditions for allocative efficiency. The regression work involves specifying two regression models with salary and student satisfaction as the respective dependent variables and teacher traits as the independent ones; estimating the equations' respective coefficients; and finally, plugging those coefficients into the corresponding marginal product to price ratios--the ratios required by microeconomics to determine allocative efficiency.

The original sample of 5,076 student-level questionnaires comes from a medium-sized public high school in the San Francisco Bay Area whose students are largely middle-class.

The 31 questionnaire items used to measure teacher traits essentially cover all achievement-producing traits noted by R.W. Heath and M.A. Nielson, Rosenshine and Furst[105] and some of those tested by Gage and Hedges.[106] Moreover, they cover most of those satisfaction-related traits revealed by the earlier-mentioned French-Lazovik[107] and Veldman and Peck[108] studies.

3.2 OBTAINING A MEASURE OF OVERALL STUDENT SATISFACTION

Overall student satisfaction with teacher performance was measured as the percentage of each teacher's students who responded "yes" to the question "Would you take a class with this teacher again?" This measure is analogous to the economists' notion of "repeat buyers," i.e., it does the best we can do in practice to ascertain who would go back for more of the same teacher, if free to do so. More specifically, though economists would much rather observe the act of repeat buying, rather than a mere response to the question "would you buy again," the rigidities of high school class scheduling generally prevent the student from ever performing that act.

3.3 SINGLE-SHEET SUMMARY OF MODEL AND METHODS

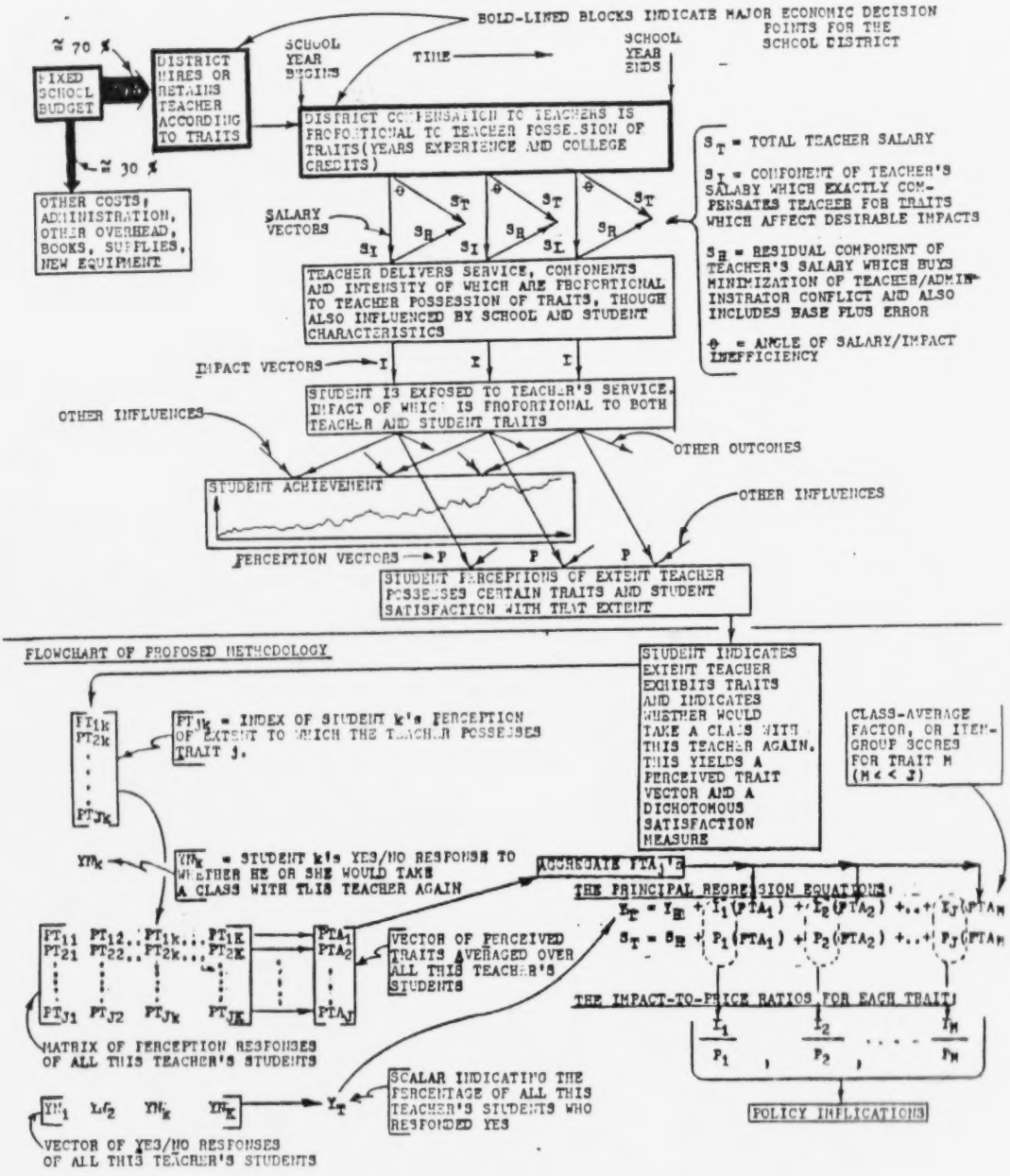
Methodologies here embody both process and input-output views of schooling. More specifically, exploring links between schooling dollars and student satisfaction begins by articulating processes that should efficiently translate budgets into teaching that attracts, challenges, and fulfills students--a process summarized on the single sheet of Figure 3.1 and described as follows:

The roughly 70 percent of district budgets that go for teachers' salaries can be disaggregated into components key to this study. Total salary (S_T, or Salary Total) appears in Figure 3.1 as a vector--the vector sum of that component that exactly compensates the teacher for traits yielding desired impact (labeled S_I for Salary yielding desired Impact) plus that residual (S_R, or Salary Residual) having nothing to do with that impact (e.g., rewards for remaining with the district). In theory, then, that part of total salary that compensates desired impact is the projection of S_I onto S_T. In turn, S_I can be viewed as the sum of manifold cost vectors, each corresponding to a specific impact-producing teacher trait.

The total impact a teacher has on students over time (shown in Figure 3.1, as "impact vectors" labeled with I's) is then proportional to levels of certain teacher traits and school and student characteristics (e.g., powerful nonteacher influences like student's socioeconomic status). This impact stimulates cognitive achievement, perceptions of teacher traits, and other outcomes (shown with arrows pointing different directions).

In short, this study explores those perceptions by having students indicate how much their teacher exhibits specific traits (indicated by the PT_{jk}'s, or Perceived Trait j evaluated by the kth student) and also whether the teaching would attract them to

FIGURE 3.1
FLOWCHART OF TEACHING SERVICE EXCHANGE,
AND THIS STUDY'S METHODOLOGIES


BOLD-LINED BLOCKS INDICATE MAJOR ECONOMIC DECISION POINTS FOR THE SCHOOL DISTRICT
≈ 70 %
FIXED SCHOOL BUDGET
DISTRICT HIRES OR RETAINS TEACHER ACCORDING TO TRAITS
SCHOOL YEAR BEGINS
TIME
SCHOOL YEAR ENDS
DISTRICT COMPENSATION TO TEACHERS IS PROPORTIONAL TO TEACHER POSSESSION OF TRAITS (YEARS EXPERIENCE AND COLLEGE CREDITS)
≈ 30 %
OTHER COSTS; ADMINISTRATION, OTHER OVERHEAD, BOOKS, SUPPLIES, NEW EQUIPMENT
SALARY VECTORS
S_T = TOTAL TEACHER SALARY
S_I = COMPONENT OF TEACHER'S SALARY WHICH EXACTLY COMPENSATES TEACHER FOR TRAITS WHICH AFFECT DESIRABLE IMPACTS
S_R = RESIDUAL COMPONENT OF TEACHER'S SALARY WHICH BUYS MINIMIZATION OF TEACHER/ADMINISTRATOR CONFLICT AND ALSO INCLUDES BASE PLUS ERROR
θ = ANGLE OF SALARY/IMPACT INEFFICIENCY
TEACHER DELIVERS SERVICE, COMPONENTS AND INTENSITY OF WHICH ARE PROPORTIONAL TO TEACHER POSSESSION OF TRAITS, THOUGH ALSO INFLUENCED BY SCHOOL AND STUDENT CHARACTERISTICS
IMPACT VECTORS
STUDENT IS EXPOSED TO TEACHER'S SERVICE, IMPACT OF WHICH IS PROPORTIONAL TO BOTH TEACHER AND STUDENT TRAITS
OTHER INFLUENCES
OTHER OUTCOMES
STUDENT ACHIEVEMENT
OTHER INFLUENCES
PERCEPTION VECTORS
STUDENT PERCEPTIONS OF EXTENT TEACHER POSSESSES CERTAIN TRAITS AND STUDENT SATISFACTION WITH TRAIT EXTENT
FLOWCHART OF PROPOSED METHODOLOGY
STUDENT INDICATES EXTENT TEACHER EXHIBITS TRAITS AND INDICATES WHETHER WOULD TAKE A CLASS WITH THIS TEACHER AGAIN. THIS YIELDS A PERCEIVED TRAIT VECTOR AND A DICHOTOMOUS SATISFACTION MEASURE
CLASS-AVERAGE FACTOR, OR ITEM-GROUP SCORES FOR TRAIT M (M << J)
PT_{1k} = INDEX OF STUDENT k's PERCEPTION OF EXTENT TO WHICH THE TEACHER POSSESSES TRAIT j.
YN_k = STUDENT k's YES/NO RESPONSE TO WHETHER HE OR SHE WOULD TAKE A CLASS WITH THIS TEACHER AGAIN
AGGREGATE PTA_j'S
THE PRINCIPAL REGRESSION EQUATIONS:
$Y_T = Y_R + I_1(PTA_1) + I_2(PTA_2) + \ldots + I_J(PTA_J)$
$S_T = S_R + P_1(PTA_1) + P_2(PTA_2) + \ldots + P_J(PTA_J)$
VECTOR OF PERCEIVED TRAITS AVERAGED OVER ALL THIS TEACHER'S STUDENTS
THE IMPACT-TO-PRICE RATIOS FOR EACH TRAIT:
$\frac{I_1}{P_1}$, $\frac{I_2}{P_2}$, $\frac{I_M}{P_M}$
POLICY IMPLICATIONS
MATRIX OF PERCEPTION RESPONSES OF ALL THIS TEACHER'S STUDENTS
SCALAR INDICATING THE PERCENTAGE OF ALL THIS TEACHER'S STUDENTS WHO RESPONDED YES
VECTOR OF YES/NO RESPONSES OF ALL THIS TEACHER'S STUDENTS

take a course with the teacher again (shown by the YN_k, a Yes or No response of the kth student). Teacher-level aggregates--Perceived Traits Averaged, shown by PTA_j's, and the percent responding "Yes" to the above teaching question (the Y_t's)--were then generated from the student-level questionnaires.

The PTA_j's served as independent variables, and Y_t's and teachers' salaries, as dependent variables for the marginal impact and marginal price regression models, respectively (as shown in the lower right hand corner of Figure 3.1, labeled "THE PRINCIPAL REGRESSION EQUATIONS"). Several different regression models (simple planar, simple quadratic, Cobb Douglas, logit and probit) helped estimate the marginal impacts of teacher traits on overall student satisfaction (components of the I vector: I_1, I_2, . . . I_M--shown under Figure 3.1's heading "THE IMPACT-TO-PRICE RATIOS FOR EACH TRAIT") and the implicit marginal prices paid for those traits (components of the P vector: P_1, P_2, . . . P_M). The simple planar equations in the lower right corner of the flow chart produced linear estimates of the raw I's and P's, which, in turn, represented the increment in student satisfaction (percent total yes) and the marginal price (the implicit marginal increase in total teacher salary, S_T), respectively, per unit change in a teacher's class-average score for trait m, ceteris paribus.

In short, computing the marginal-impact-to-marginal-price (I/P) ratio for each trait produced disparate I/P ratios across traits. This, in turn, suggested specific inefficiencies and corresponding policy implications.

4.1 OVERVIEW OF HOW THE VARIABLES INTERRELATE: THE MANAGERIAL GRID FOR TEACHERS

Perhaps the best overview of how all the variables interrelate is seen in the following SPSS-X[109] quartimax[110] rotation in Table 4.1:

TABLE 4.1
QUARTIMAX ROTATION, FACTOR MATRIX USING
ALL TEACHER-LEVEL VARIABLES, BOTH YEARS COMBINED

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
	-----STUDENT-CARING (THIS STUDY)			
	|	PEOPLE-CENTERED (BLAKE AND MOUTON)		
UNDRST	.92764	CONSIDERATION (FLEISHMAN, ET AL.)		
INVOLV	.88818	SOCIO-EMOTIONAL SPECIALIST (BALES)		
RESPEC	.88715	RELATIONSHIP MOTIVATED (FIEDLER)		
FLEXBL	.88491	SOCIAL EXPRESSIVITY (COUCH)		
CLEARX	.86691			
WARM	.85078			
YRCONT	.84660			
UNBIAS	.84174			
PERSON	.83669			
CREATV	.83627			
OPIN	.80619			
COMMUN	.79943			
IMPORT	.78704			
PYES	.78086			
AMNTWK	.76820			
FAIR	.76208			
ADMIT	.75304			
ENTHUS	.71907			
METHOD	.67314			
RELEV	.59898			
INFORM	.55147			
CONTRL	.55105			
PRINTSB	.51459		-----TASK-DRIVEN (THIS STUDY)	
		|	PRODUCTION-CENTERED	
PREPAR		.90605	(BLAKE AND MOUTON)	
ORGNZD		.87681	INITIATING STRUCTURE	
CTIME		.82032	(FLEISHMAN, ET AL.)	
KNOWL		.78549	TASK SPECIALIST (BALES)	
MISSED		-.63689	TASK MOTIVATED (FIEDLER)	
AVAIL		.58244	TASK SERIOUSNESS (COUCH)	
RETURNS		.57769		
CRITIC	.51694	.54117		

TABLE 4.1, CONTINUED
QUARTIMAX ROTATION, FACTOR MATRIX USING
ALL TEACHER-LEVEL VARIABLES, BOTH YEARS COMBINED

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
PHWOCCA			-.95045	STUDENTS'
PHWALWS			.95045	WORK AND
HMWORK			-.92664	EVALUATION
LGRADE			.61684	
TESTS			.51948	
PHADTO			REASON FOR	-.97378
PROTHER			TAKING THE	.96245
PRTEACH			COURSE	-.90992

	FACTOR 5	FACTOR 6
SALARY	.91782	SALARY
YEARS	.87422	AND ITS
EDUNITS	.80756	DETERMINANTS
PBLKHIS		.95620 ETHNICITY
PHISP		.89316 MIX IN CLASS
PWHITE		-.82414

Though the questionnaire was never consciously designed to produce specific factors, the above loadings suggest variable groupings with the follow descriptions:

FACTOR 1: The extent to which the teacher cares about the socio-emotional well-being of students (STUDENT-CARING).

FACTOR 2: The degree to which the teacher is driven toward completion of the tasks at hand (TASK-DRIVEN);

FACTOR 3: The student's work and evaluation;

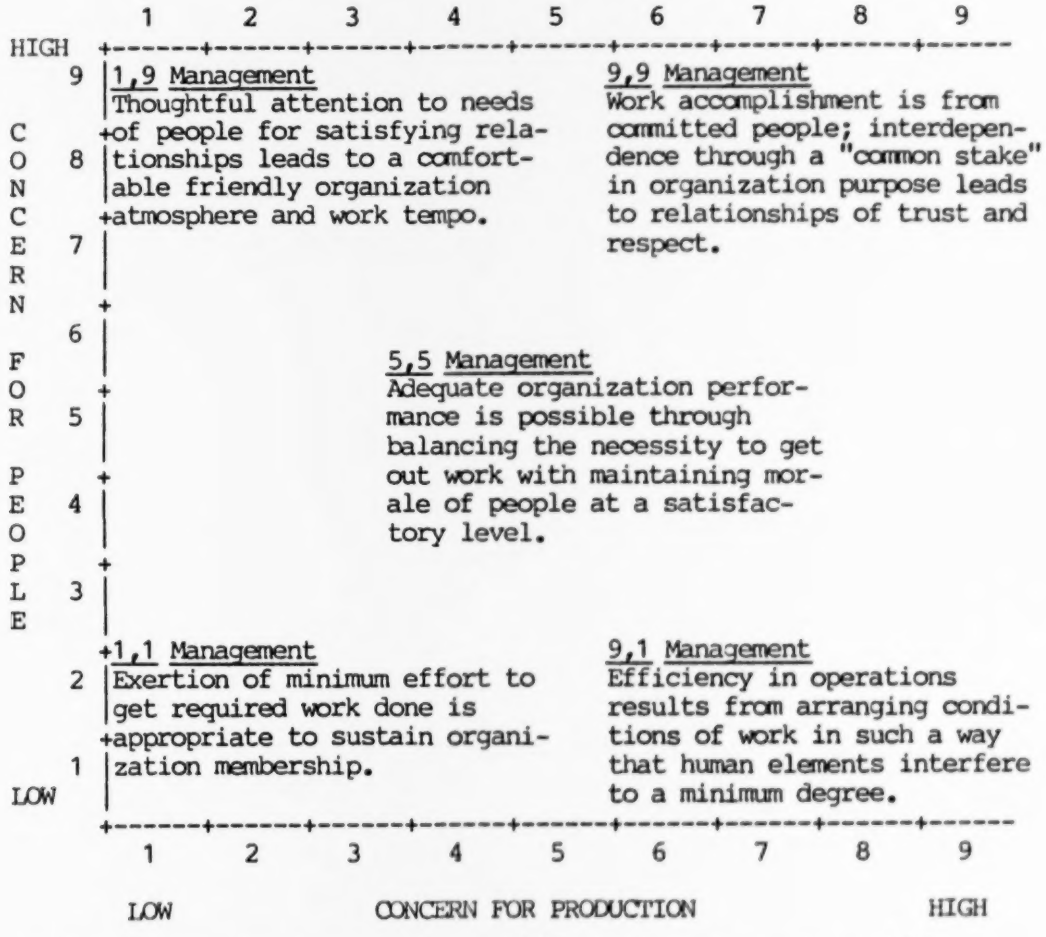
FACTOR 4: The student's reasons for taking the course;

FACTOR 5: The teacher's salary variables; and,

FACTOR 6: The ethnicity mix of the students within a given teacher's aggregation;

The first two factors, STUDENT-CARING and TASK-DRIVEN, account for 69.6 percent of the total variation among teacher trait variables alone[11] and replicate the same two latent dimensions that persist across other studies from psychology and

FIGURE 4.1
THE MANAGERIAL GRID OF BLAKE AND MOUTON[112]


1 2 3 4 5 6 7 8 9
HIGH
9
1,9 Management
Thoughtful attention to needs of people for satisfying relationships leads to a comfortable friendly organization atmosphere and work tempo.
9,9 Management
Work accomplishment is from committed people; interdependence through a "common stake" in organization purpose leads to relationships of trust and respect.
CONCERN FOR PEOPLE
8
7
6
5,5 Management
Adequate organization performance is possible through balancing the necessity to get out work with maintaining morale of people at a satisfactory level.
5
4
3
1,1 Management
Exertion of minimum effort to get required work done is appropriate to sustain organization membership.
2
1
LOW
9,1 Management
Efficiency in operations results from arranging conditions of work in such a way that human elements interfere to a minimum degree.
1 2 3 4 5 6 7 8 9
LOW
CONCERN FOR PRODUCTION
HIGH

managerial leadership (e.g., Bales,[113] Blake and Mouton,[114] Couch,[115] Fiedler,[116] and Fleishman et al.[117]). Labeling these dimensions as "concern for people" and "concern for production" Blake and Mouton plot and discuss five managerial styles on their resulting "Managerial Grid," reproduced in Figure 4.1. According to Blake and Mouton, that grid:

> . . . has been applied in widely different organizational settings in the United States, Canada, Europe and Asia. Included are industrial facilities of manufacturing, sales, R&D, and union organizations as well as military, governmental, professional and welfare settings such as community agencies . . . it seems to provide descriptions of managerial alternatives that are equally useful when applied to managerial dilemmas in the U.S., in countries of Europe that are somewhat similar to our own, and in

> cultures of Asia, which are far different. The schema, in other words, seems to be relatively culture-free and, therefore, of general relevance for understanding problems of management wherever men work in concert . . . its application is not limited to any particular level in the organization hierarchy. It applies in solving problems at the bottom where concrete supervisory skills are required and at the top where executive decision-making involving far more subtle and complex judgments is demanded.[118]

Likewise, Fleishman, Harris and Burtt describe what appear to be the same factors discussed by Blake and Mouton:

> Actually, we were able to extract two leadership factors which were quite independent, both of which will be discussed in detail later. One identified as 'consideration' included such characteristics between supervisors and subordinates as friendliness, mutual trust, and respect. The other, termed 'initiating structure,' involved establishing patterns of organization, channels of communication, and ways of getting things done.[119]

The four teacher traits loading heaviest on the "student caring" versus "task driven" factors of this study's "Managerial Grid for Teachers," and the other authors' respective versions of its dimensions, appear in Table 4.2.

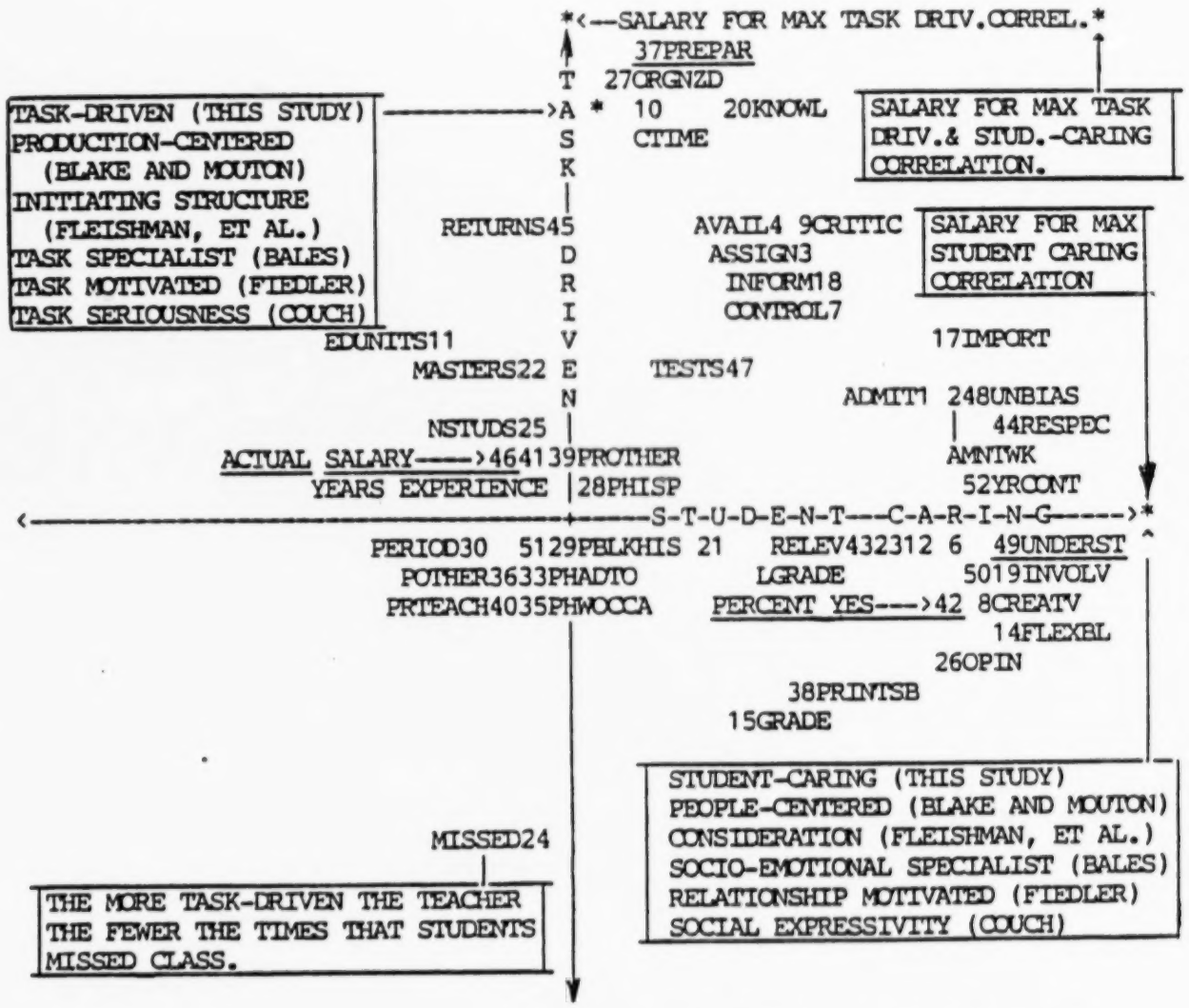
TABLE 4.2
FOUR VARIABLES LOADING HEAVIEST ON STUDENT-CARING AND TASK-DRIVEN FACTORS AND OTHER AUTHOR'S LABELS FOR THEIR VERSIONS OF THESE FACTORS

FACTOR LABELS FOR THIS STUDY	STUDENT-CARING	TASK DRIVEN
VARIABLES LOADING HEAVIEST ON THESE FACTORS	DEGREE TO WHICH YOUR TEACHER: 1. IS UNDERSTANDING (UNDRST) 2. STIMULATES YOUR INTEREST AND INVOLVEMENT IN THE COURSE (INVOLV) 3. IS RESPECTED BY YOU (RESPEC) 4. IS FLEXIBLE (FLEXBL)	DEGREE TO WHICH YOUR TEACHER: 1. IS PREPARED FOR CLASS (PREPAR) 2. IS WELL ORGANIZED (ORGNZD) 3. MAKES CONSTRUCTIVE USE OF CLASS TIME (CTIME) 4. IS KNOWLEDGEABLE ABOUT THE SUBJECT MATTER (KNOWL)
AUTHORS' LABELS FOR SIMILAR FACTORS FOUND IN PSYCHOLOGY AND MANAGEMENT LITERATURE:		
BALES	SOCIO-EMOTIONAL SPECIALIST	TASK SPECIALIST
BLAKE AND MOUTON	PEOPLE-CENTERED	PRODUCTION-CENTERED
COUCH	SOCIAL EXPRESSIVITY	TASK SERIOUSNESS
FIEDLER	RELATIONSHIP MOTIVATED	TASK MOTIVATED
FLEISHMAN, ET AL.	CONSIDERATION	INITIATING STRUCTURE

The other authors' labels for their versions of these two dimensions have also been inserted in the following plot of the factor loadings (Figure 4.3).

It should be no surprise that evaluations of teachers produce the very two dimensions widely believed to be key among managers. Though big differences exist between their respective jobs, both are responsible for getting a subordinate group of people to accomplish prespecified tasks, and reward them for it.

FIGURE 4.3
QUARTIMAX ROTATION PLOT OF FIRST TWO OF THE ABOVE FACTORS:
A MANAGERIAL GRID FOR TEACHERS



Note: The overall satisfaction measure, PERCENT YES, loads on TASK-DRIVEN negatively here, but positively in other rotation methods. For the main analyses, however, the varible chosen to represent TASK DRIVEN, "IS PREPARED FOR CLASS" (variable 37 above) does relate positively to satisfaction.

4.2 THE GRID AND RELATIONSHIPS AMONG KEY VARIABLES

The meanings of other variables plotted in Figure 4.3 also seem to agree with their locations within the TASK-DRIVEN versus STUDENT-CARING grid. For example, the strongly negative loading of "how many class periods of this course have you missed" (MISSED24 on the plot) confirms that the more task-driven the teacher, the fewer the class periods the students will miss (or will report having missed).

Of considerably greater importance, here, is that overall satisfaction (the percentage of students who want to take a course again with the teacher, PERCENT YES--->42 on the plot) loads quite positively on STUDENT-CARING, and, ambiguously on TASK-DRIVEN (though its positive correlation with the strongest proxy for task-driven, is less unambiguous).[120] Moreover, its loadings are distant from, and carry opposite signs of, the loadings for salary. Taken together, this suggests that student satisfaction relates strongly to student-caring, considerably less to task-driven, and perhaps negatively, if at all, to salary. In short, this hints that salary is not allocated to produce satisfaction efficiently.

5. SPECIFYING, AND ESTIMATING THE EQUATIONS

Despite distinct differences in each rotation method's objectives, Table 5.1 shows that for this particular sample (both years of teacher-level aggregates combined and weighted) each produced roughly similar rank-orderings of variables loading high on each of both factors. Indeed, the consistently high loadings of the same variables across the five different techniques argues for using these variables as proxies for the grid dimensions they load high on. Substituting proxies for the grid dimensions will, in turn, avoid what in earlier runs involved messy computations and interpretations of factor scores; a substitution consistent with our goal of constructing a model that is parsimonious, and easy to interpret.

Table 5.1 shows that the STUDENT-CARING versus TASK-DRIVEN factors will be best represented by either UNDERSTANDING and ORGANIZED (ranked first on each of the two dimensions for quartimax) or UNDERSTANDING and PREPARED (first ranked for both varimax and equamax).

TABLE 5.1
RANK ORDERING OF FOUR HIGHEST LOADING VARIABLES ON
STUDENT-CARING VERSUS TASK-DRIVEN FACTORS
FOR DIFFERENT ROTATION TECHNIQUES

	QUARTIMAX	VARIMAX	EQUAMAX	OBLIMIN
STUDENT-CARING:				
FIRST	UNDERSTANDING	UNDERSTANDING	UNDERSTANDING	FLEXIBLE
SECOND	RESPECTED	FLEXIBLE	FLEXIBLE	UNDERSTANDING
THIRD	PERSON	YOURCONTRIB	YOURCONTRIB	YOURCONTRIB
FOURTH	YOURCONTRIB	WARM	WARM	WARM
TASK-DRIVEN:				
FIRST	ORGANIZED	PREPARED	PREPARED	ORGANIZED
SECOND	PREPARED	ORGANIZED	ORGANIZED	PREPARED
THIRD	CLASSTIME	CLASSTIME	CLASSTIME	CLASSTIME
FOURTH	KNOWLEDGE	KNOWLEDGE	KNOWLEDGE	KNOWLEDGE

Where, in alphabetical order, the above variable names correspond to the following questionnaire items, measuring the extent to which the teacher exhibits the indicated trait:

CLASSTIME:	Makes constructive use of class time.
FLEXIBLE:	Is flexible.
INVOLVED:	Stimulates your interest and involvement in the course.
KNOWLEDGE:	Is knowledgeable about the subject matter.
ORGANIZED:	Is well organized.
PERSON:	Respects you as an individual; seems to be interested in you as a person.
PREPARED:	Is prepared for class.
RESPECTED:	Is respected by you.
UNDERSTANDING:	Is understanding.
WARM:	Is warm/friendly.
YOURCONTRIB:	Makes you feel that your contributions are important and meaningful.

5.1 ESTIMATING THE COEFFICIENTS

After numerous attempts with more complicated models, the preferred models contain only the earlier determined proxies for each of the grid dimensions, and the inputs the district actually pays for (teacher's post-B.A. education units, and teacher's years experience), for a total of four independent variables.

Regressions were done using five models: simple planar, simple quadratic, Cobb-Douglas, probit, and logit. In short, the simple planar model has PYES move as a linearly increasing function of a given trait variable (ceteris paribus), while the quadratic has PYES move as a curvilinearly increasing (squared) function--with decreasing slope (diminishing returns). Though Cobb-Douglas, probit and logit are also curvilinear, Cobb-Douglas embodies a slightly more sophisticated exponential of which the natural log is taken; Probit involves the cumulative normal

TABLE 5.2
ADJUSTED R-SQUARED VALUES FOR THE DIFFERENT MODELS
RUN USING SPSS-X'S WEIGHTING METHOD

	STUDENT CARING PROXY--->UNDERSTANDING TASK-DRIVEN PROXY------>ORGANIZED	UNDERSTANDING PREPARED
1. SIMPLE PLANAR		
A. DIRECT	0.536 (RUN 3)	0.536 (RUN 11)
B. CONTROLLED	0.408 (RUN 4)	0.409 (RUN 12)
2. SIMPLE QUADRATIC		
A. DIRECT	0.609 (RUN 5)	0.627 (RUN 13)
B. CONTROLLED	0.490 (RUN 6)	0.506 (RUN 14)
3. COBB-DOUGLAS		
A. DIRECT	0.346 (RUN 7)	0.331 (RUN 15)
B. CONTROLLED	0.406 (RUN 8)	0.371 (RUN 16)
4. PROBIT		
A. DIRECT	0.422 (RUN 9)	0.417 (RUN 17)
B. CONTROLLED	NOT RUN	NOT RUN
5. LOGIT		
A. DIRECT	0.479 (RUN 10)	0.475 (RUN 18)
B. CONTROLLED	NOT RUN	NOT RUN

6. SIMPLE PLANAR REGRESSION INCLUDING THE ACTUAL PAID-FOR INPUTS, TEACHERS POST-B.A. UNITS, AND TEACHER'S YEARS EXPERIENCE.

	STUDENT CARING PROXY--->UNDERSTANDING TASK-DRIVEN PROXY------>ORGANIZED	FLEXIBLE PREPARED
A. DIRECT	0.570 (RUN 1)	0.570 (RUN 2)
B. CONTROLLED	0.451 (RUN 3)	0.453 (RUN 4)
C. SALARY EQUATIONS	0.951 (RUN 5)	0.954 (RUN 6)

probability function; And logit uses a "logistic function" of the odds that a given student will respond "yes" to the question "Would you take a class again with this teacher?"

Table 5.2's rows labeled "DIRECT" indicate where PYES was used directly as the dependent student satisfaction variable, while rows labeled "CONTROLLED," refer to regressions that <u>controlled</u> PYES for effects beyond the teacher's control (effects discovered earlier, which we had promised to deal with at this regression stage).

Despite the virtues of "controlled" regressions over "direct" ones, direct versions were still run to understand the nature of the raw production function, and to see the degree to which the controlling process "dulls" the regression equations in terms of lower R-squareds, wider standard errors about the coefficients, and lower significance levels. As the adjusted R-squareds of Table 5.2 show, this dulling effect is considerable.

5.3 SIMPLE PLANAR APPROACH

Consistent with the hope that the model be parsimonious, the earlier-mentioned trait proxies for STUDENT CARING (UNDRST) AND TASK DRIVEN (PREPAR) were plugged into the following planar model:

$$PYES = B_0 + B_1 \text{ (STUDENT CARING)} + B_2 \text{ (TASK DRIVEN)}$$

Table 5.3 shows the resulting coefficients:

TABLE 5.3
COEFFICIENT RESULTS: SIMPLE AND RESIDUAL PLANAR MODELS

RUN 11 SIMPLE PLANAR REGRESSION OF: PYES = F(UNDRST, PREPAR)
ADJUSTED R-SQUARED = 0.536

INDEP. VARIABLE	SLOPE B	STD ERR B	T	SIG T
PREPAR	5.22761	2.01345	2.596	.0097
UNDRST	31.08833	1.45772	21.327	.0000
(Y-INTERCEPT)	-111.58839	11.29216	-9.882	.0000

Though the planar fits offer respectable R-squareds, the linearity embodied in the above will not reveal the extent to which diminishing marginal returns (curvilinear phenomena) might operate.

5.4 SIMPLE CURVILINEAR (QUADRATIC) APPROACH

To detect and measure marginal returns, the following quadratic form was run on both pairs of factor proxies:

$$PYES = B_0 + B_1 \text{ (STUDENT CARING)} + B_2 \text{ (STUDENT CARING)}^2 + B_3 \text{ (TASK DRIVEN)} + B_4 \text{ (TASK DRIVEN)}^2$$

As Table 5.2 reveals, this particular model explained more variation in satisfaction than did any other, in both direct, and controlled versions. Moreover, as Table 5.4 shows, it consistently yielded negative coefficients on the squared terms--i.e., empirical evidence the diminishing returns discussed above.

TABLE 5.4
COEFFICIENT RESULTS
FOR THE SIMPLE QUADRATIC MODEL

RUN 13 SIMPLE QUADRATIC REGRESSION OF:
PYES = F(UNDRST, PREPAR, UNDRST^2, PREPAR^2)

ADJUSTED R-SQUARED = 0.627

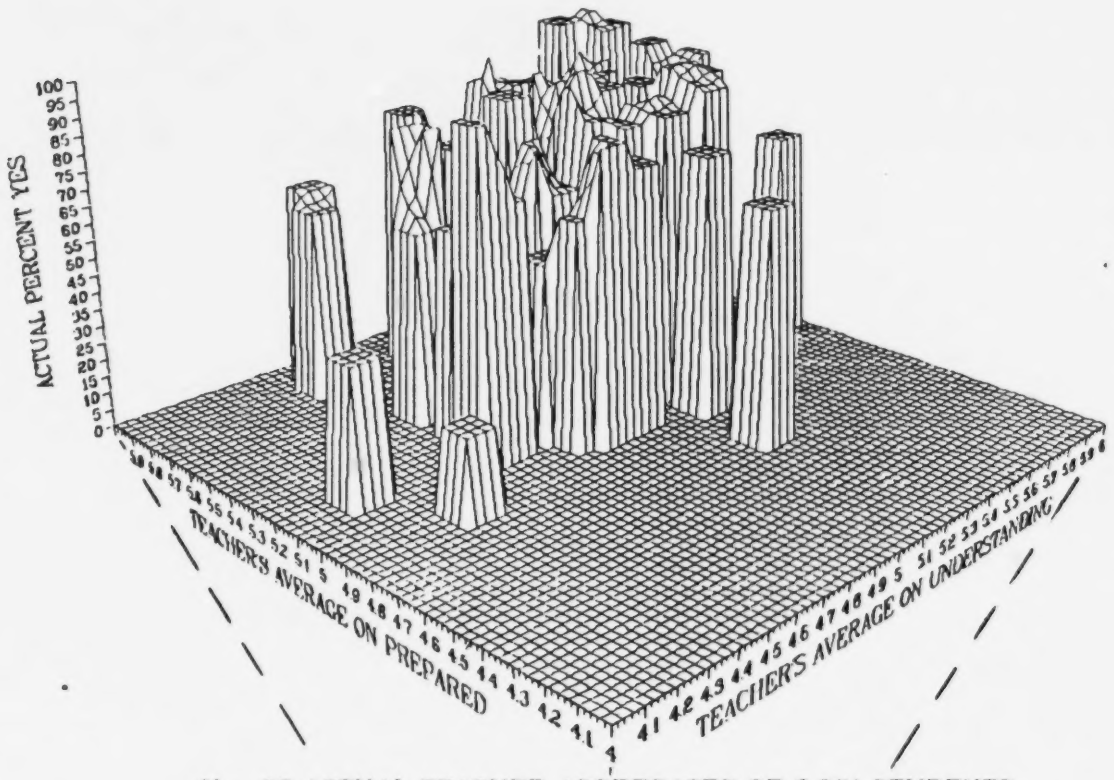
INDEP. VARIABLE	SLOPE B	STD ERR B	T	SIG T
(Y-INTERCEPT)	-1619.28693	176.32957	-9.183	.0000
UNDRST	280.19966	29.58726	9.470	.0000
PREPAR	327.74262	64.15070	5.109	.0000
UNDRST^2	-24.53956	2.91113	-8.430	.0000
PREPAR^2	-29.53668	5.91444	-4.994	.0000

5.5 CLOSER INSPECTION OF PLANAR AND QUADRATIC FORMS

Since both the planar and quadratic forms require no translation of any of the original variables (unlike Cobb-Douglas, logit, and probit), both can be discussed in the context of three-dimensional plots whose dimensions measure the locations of teacher aggregates in the relevant production space. That space will measure student satisfaction (PERCENT YES) as a function of teaching that is student-caring (proxied by UNDERSTANDING), and task-driven (proxied by PREPARED).[121] The 60 actual teacher aggregates plotted[122] in that space appear in Figure 5.2:

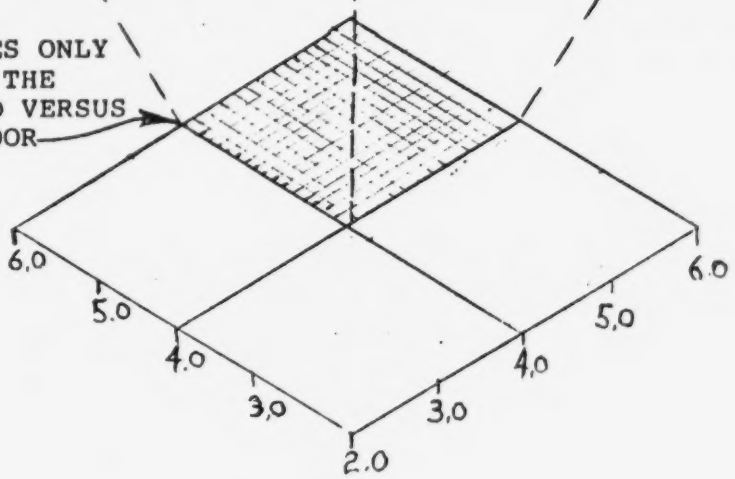
FIGURE 5.2

ACTUAL PERCENT YES AS A FUNCTION OF UNDERSTANDING VERSUS PREPARED



N = 60 ACTUAL TEACHER AGGREGATES OF 3,961 STUDENTS.

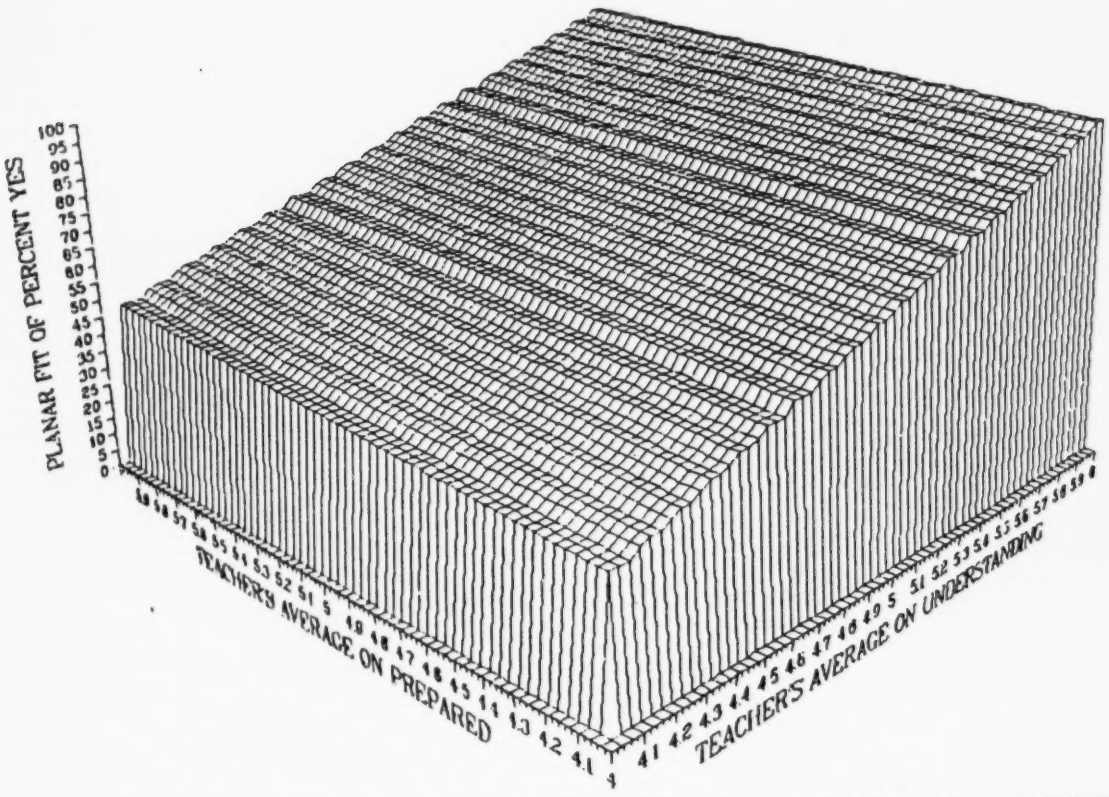
ABOVE PLOT OCCUPIES ONLY UPPER QUADRANT OF THE ORIGINAL "PREPARED VERSUS UNDERSTANDING" FLOOR



As the above plot confirms, the high ratings received on average by all teachers, confine the data to the greater-than-4.0 quadrant of the original floor. All 60 data points do not appear in the plot because the plotting software, designed to show the corresponding surface, depicts only the approximate view that would result from draping a sheet over the points. That is, it conceals all data points that lie under the sheet (All the actual data points are more conveniently plotted and discernible in subsequent figures depicting two-dimensional slices of the production space). Despite the concealment, the surface does suggest that PERCENT YES tends to rise as a function of the teacher's average for PREPARED, and UNDERSTANDING. Plots of the corresponding planar and quadratic fits (whose statistics were detailed earlier) confirm the nature of that trend (Figures 5.3 and 5.4). The four-term quadratic equation produces the expected surface with positive first, and negative second, derivatives with respect to both inputs.

FIGURE 5.3

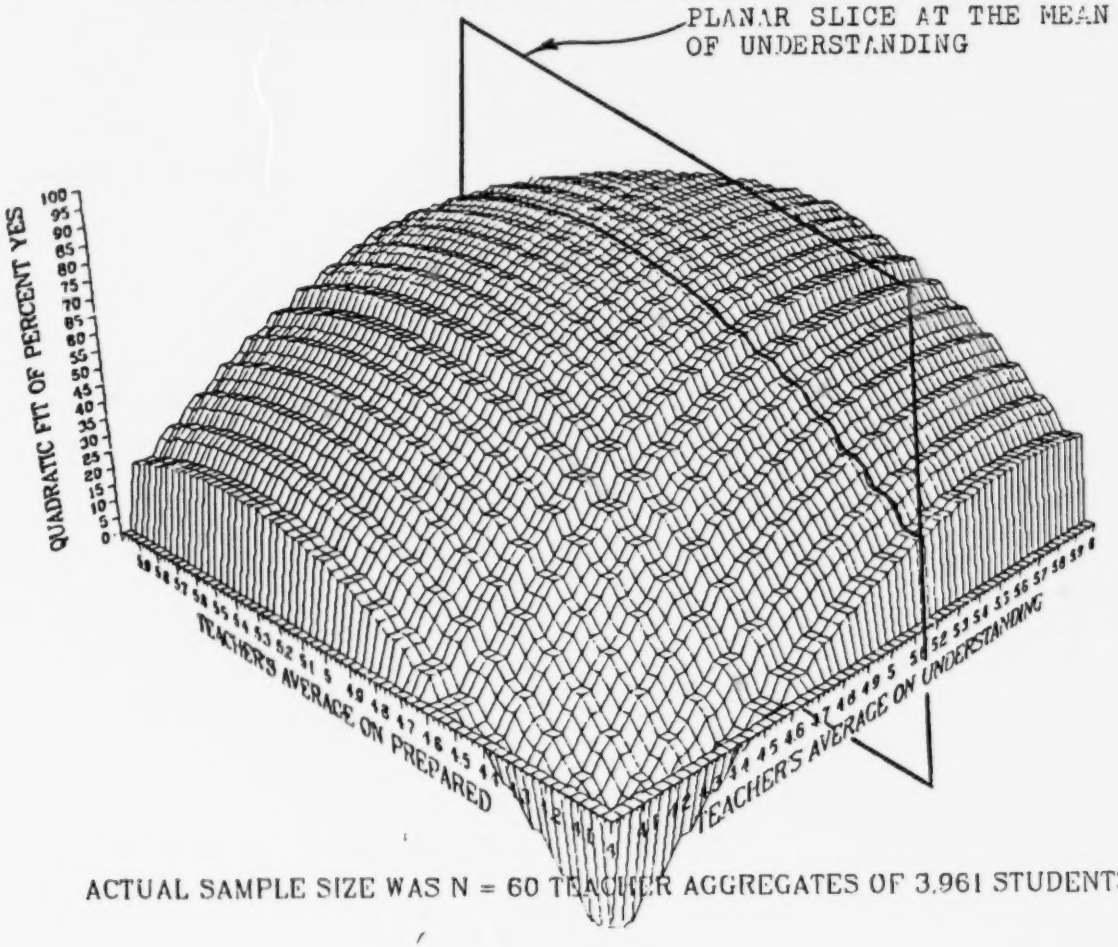
PERCENT YES AS PLANAR FUNCTION OF UNDERSTANDING VERSUS PREPARED



ACTUAL SAMPLE SIZE WAS N = 60 TEACHER AGGREGATES OF 3,961 STUDENTS.

FIGURE 5.4

PERCENT YES AS A QUADRATIC FUNCTION OF UNDERSTANDING VERSUS PREPARED


PLANAR SLICE AT THE MEAN OF UNDERSTANDING
QUADRATIC FIT OF PERCENT YES
TEACHERS AVERAGE ON PREPARED
TEACHERS AVERAGE ON UNDERSTANDING

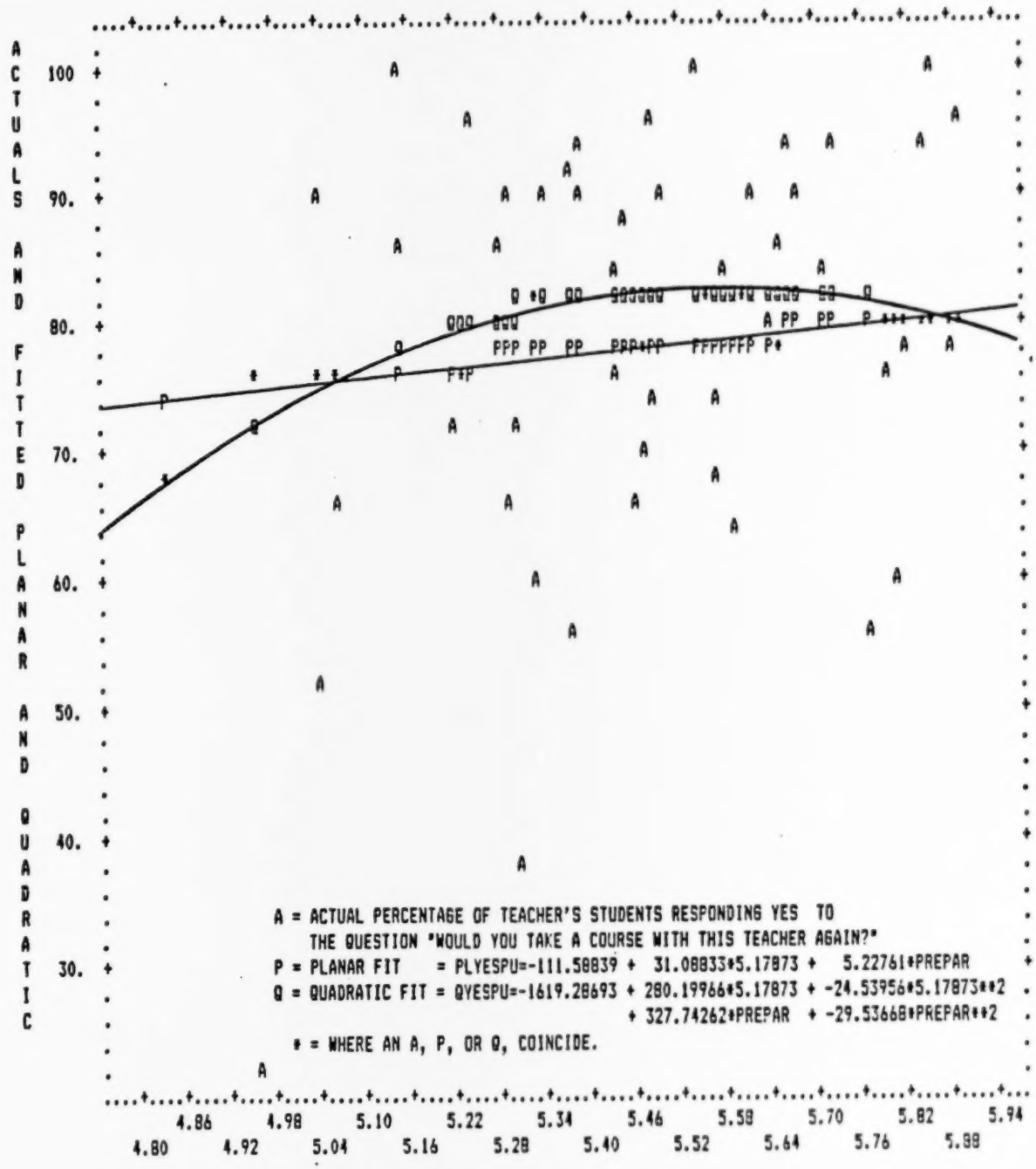
ACTUAL SAMPLE SIZE WAS N = 60 TEACHER AGGREGATES OF 3,961 STUDENTS

Detailed comparison of the planar versus quadratic fits is enhanced by taking vertical slices of the surfaces, at the means of the respective inputs. For example, Figure 5.5 represents an enhanced version of the slice shown in the three dimensional quadratic plot of Figure 5.4. This PERCENT YES versus PREPARED cross section taken at the average value for UNDERSTANDING (equal to 5.17) shows both the corresponding planar and quadratic curves, and superimposes projections of all the actual data points (represented with "A"'s for Actuals or "*"'s for where two or more points overlap). The word projections is emphasized because if one misperceives the points and curves to share the same plane, then one understates the goodness of fit between the data cloud and the fitted surfaces.

Figure 5.6 shows a second slice, this time taken at the mean of PREPARED (equal to 5.48). Together, Figures 5.5 and 5.6 reveal, as do the equations in the bottom of each plot, that student satisfaction increases most steeply with respect to student-caring, and very gradually with respect to task-driven.

FIGURE 5.5

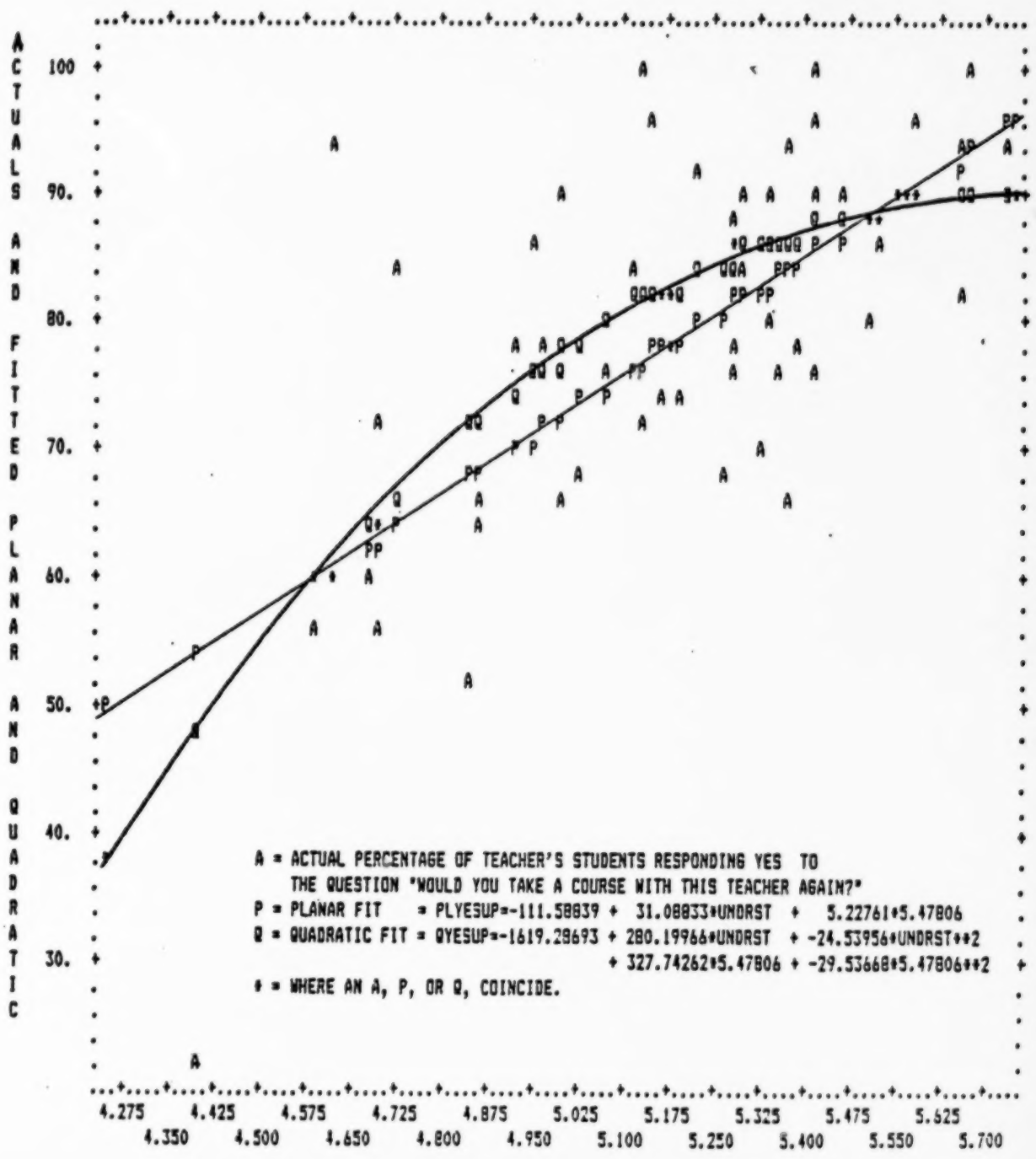
SLICE DEPICTING ACTUAL, PLANAR, AND QUADRATIC FITS PROJECTED AGAINST PYES VERSUS PREPAR PLANE, WITH UNDRSI HELD CONSTANT AT ITS MEAN. SEE EQUATION DETAIL IN LOWER PORTION OF BOX.


ACTUALS AND FITTED PLANAR AND QUADRATIC
100
90.
80.
70.
60.
50.
40.
30.
A = ACTUAL PERCENTAGE OF TEACHER'S STUDENTS RESPONDING YES TO THE QUESTION "WOULD YOU TAKE A COURSE WITH THIS TEACHER AGAIN?"
P = PLANAR FIT = PLYESPU=-111.58839 + 31.08833*5.17873 + 5.22761*PREPAR
Q = QUADRATIC FIT = QYESPU=-1619.28693 + 280.19966*5.17873 + -24.53956*5.17873**2 + 327.74262*PREPAR + -29.53668*PREPAR**2
* = WHERE AN A, P, OR Q, COINCIDE.
4.80 4.86 4.92 4.98 5.04 5.10 5.16 5.22 5.28 5.34 5.40 5.46 5.52 5.58 5.64 5.70 5.76 5.82 5.88 5.94

X = TASK-DRIVEN PROXY = PREPAR = TEACHER'S AVERAGE RATING ON: "THE DEGREE TO WHICH YOU THINK YOUR TEACHER IS PREPARED FOR CLASS."

FIGURE 5.6

SLICE DEPICTING ACTUAL, PLANAR, AND QUADRATIC FITS PROJECTED AGAINST PYES VERSUS UNDRST PLANE, WITH PREPAR HELD CONSTANT AT ITS MEAN. SEE EQUATION DETAIL IN LOWER PORTION OF BOX.


ACTUALS AND FITTED PLANAR AND QUADRATIC
100
90.
80.
70.
60.
50.
40.
30.
A = ACTUAL PERCENTAGE OF TEACHER'S STUDENTS RESPONDING YES TO THE QUESTION "WOULD YOU TAKE A COURSE WITH THIS TEACHER AGAIN?"
P = PLANAR FIT = PLYESUP=-111.58839 + 31.08833*UNDRST + 5.22761*5.47806
Q = QUADRATIC FIT = QYESUP=-1619.28693 + 280.19966*UNDRST + -24.53956*UNDRST**2
+ 327.74262*5.47806 + -29.53668*5.47806**2
+ = WHERE AN A, P, OR Q, COINCIDE.
4.275 4.425 4.575 4.725 4.875 5.025 5.175 5.325 5.475 5.625
4.350 4.500 4.650 4.800 4.950 5.100 5.250 5.400 5.550 5.700

I = STUDENT-CARING PROXY = UNDRST = TEACHER'S AVERAGE RATING ON:
"DEGREE TO WHICH YOU THINK YOUR TEACHER IS UNDERSTANDING"

Prior to the above plots, the quadratic model appeared to offer several substantial advantages over the planar form. It captured the most variation of all the models while producing significant coefficients that reflect diminishing returns and allow for determination of point marginal products. Yet as the slices show, the point marginal products (slopes) are sensitive to the particular grid coordinates at which those surface points are evaluated. By contrast, the planar fit reflects a good approximation to the quadratic one, and embodies what might be seen as an "overall average" slope that is constant and positive throughout the range of the respective input. Though the quadratic's point-specific slopes are more accurate than the planar form's "average" view, there is no stringent need here for such accuracy. The planar form offers more simplicity, and, slope values generalizable to the entire range of input values--while closely approximating the corresponding quadratic slopes. Moreover, the simpler planar model will greatly aid the final process that must include estimating marginal products and marginal prices of not only UNDERSTANDING and PREPARED, but also the inputs for which the district pays explicit prices (years experience and educational units). Finally the reader is reminded that this is the most common form of educational production function. For the above reasons, the final regressions embodied the following simple planar form:

$$\underset{\text{PYES}}{\text{PYES}} = \text{BASE} + MI_1(TT_1) + MI_2(TT_2) + \ldots + MI_K(TT_K)$$

5.6 FINAL REGRESSIONS

Table 5.3 presented coefficients for the planar form using just the two grid-dimension proxies as predictors. However, to analyze the school district's actual allocation policy relative to those two theory-based inputs, the inputs to which the district actually allocates (teacher's years experience, coded "YEARS," and education beyond the bachelor's degree, coded "EDUNITS") must also be included.[123]

To estimate the marginal impact coefficients, YEARS and EDUNITS were added to the equation predicting PYES. The resulting coefficients appear in Table 5.5:

TABLE 5.5
SYSTAT WEIGHTED MODEL WITHOUT INTERCEPT
DEPENDENT VARIABLE = PYES

INDEP. VARIABLE	SLOPE B	STD ERR B	T	SIG T
UNDRST	33.952	3.905	-4.02	.000
PREPAR	5.015	5.376	0.93	.350
YEARS	-1.256	0.633	-1.99	.050
EDUNITS	0.087	0.104	0.83	.400

5.7 FINAL RATIOS

With the marginal impact estimates in hand, calculating the following marginal-impact-to-marginal-price ratios,

$$\frac{MI_1}{MP_1} \;?\; \frac{MI_2}{MP_2} \;?\; \ldots \;?\; \frac{MI_K}{MP_K}$$

requires that we now estimate the implicit marginal prices (the MP_i's of the denominators). These are obtained by replacing only PYES with SALARY in the same model that generated the MI_i's, then rerunning the regressions; i.e., by changing:

$$\underset{\text{PYES}}{PYES} = \underset{\text{BASE}}{BASE} + MI_1(TT_1) + MI_2(TT_2) + \ldots + MI_K(TT_K)$$

to:

$$\underset{\text{SAL}}{SALARY} = \underset{\text{BASE}}{BASE} + MP_1(TT_1) + MP_2(TT_2) + \ldots + MP_K(TT_K)$$

and estimating the new coefficients.

The coefficients from the student-weighted regressions predicting satisfaction, combined with those from the unweighted regressions predicting salary, yield the long-sought "marginal impact to marginal price" ratios of Table 5.6.

The extreme disparities among Table 5.6's quotients, suggest that the district allocates its budget inefficiently with respect to producing teaching that attracts students. The quotients

TABLE 5.6
FINAL MARGINAL IMPACT TO MARGINAL PRICE RATIOS
AND RELATED STATISTICAL DETAIL
(* => $p \leq .05$)

PREDICTORS--->	UNDRST= IS UNDER-STANDING	PREPAR= IS PREPARED FOR CLASS	YEARS= YEARS TEACHING EXPERIENCE	UNITS= EDUCATION UNITS BEYOND B.A.
(STD. ERROR B)	(3.905)	(5.376)	(0.633)	(0.104)
(T STATISTIC)	(-4.02)	(0.93)	(-1.99)	(0.83)
(SIG. OF COEFS.)	(.000)	(.355)	(.052)	(.408)
(PART. CORR.)	(.73)	(.30)	(-.19)	(-.17)
ACTUAL COEFFICIENTS OF PYES EQUATION------>	SLOPES FROM SYSTAT WEIGHTED NO-CONSTANT REGRESSION ADJUSTED MULTIPLE R-SQUARED = .430			
	33.952	5.015	-1.256	0.087
	--------- ≠	-------- ≠	---------- ≠	--------
ACTUAL------>	-169.057	*819.285*	*811.026*	*25.347*
COEFFICIENTS OF SALARY EQUATION	SLOPES FROM 60 TEACHER UNWEIGHTED SALARY REGRESSION ADJUSTED MULTIPLE R-SQUARED = .960			
(STD. ERROR B)	(300.664)	(50.357)	(432.621)	(8.242)
(T STATISTIC)	(-.562)	(16.269)	(1.874)	(3.075)
(SIG. OF COEFS.)	(.5000)	(.0001)	(.0500)	(.0030)
(PART. CORR.)	(.075)	(.909)	(.245)	(.383)
MI/MP QUOTIENTS-->	-0.201	≠ 0.006	≠ -0.002	≠ 0.003

TABLE 7.7
CORRELATION MATRIX OF MAIN ANALYSIS VARIABLES
USED TO OBTAIN THE ABOVE RATIOS

The following was computed from teacher-level aggregates where each aggregate was weighted by the square root of the number of students in the aggregate (* => $p \leq .05$).

	EDUNITS	PREPAR	PYES	SAL	UNDRST	YEARS
EDUNITS	1.00	.21	-.17	.86*	-.10	.82*
PREPAR		1.00	.30*	.19	.31*	.12
PYES			1.00	-.20	.73*	-.19
SAL				1.00	-.04	.97*
UNDRST					1.00	-.03
YEARS						1.00

themselves, however, cloud the more useful picture offered by their marginal-impact and marginal-product numerators and denominators. That picture begins with TABLE 5.8's discussion of the coefficients:

TABLE 5.8
DISCUSSION OF THE PRECEDING
MARGINAL-IMPACT-TO-MARGINAL-PRICE COEFFICIENTS
(* => $p \leq .05$)

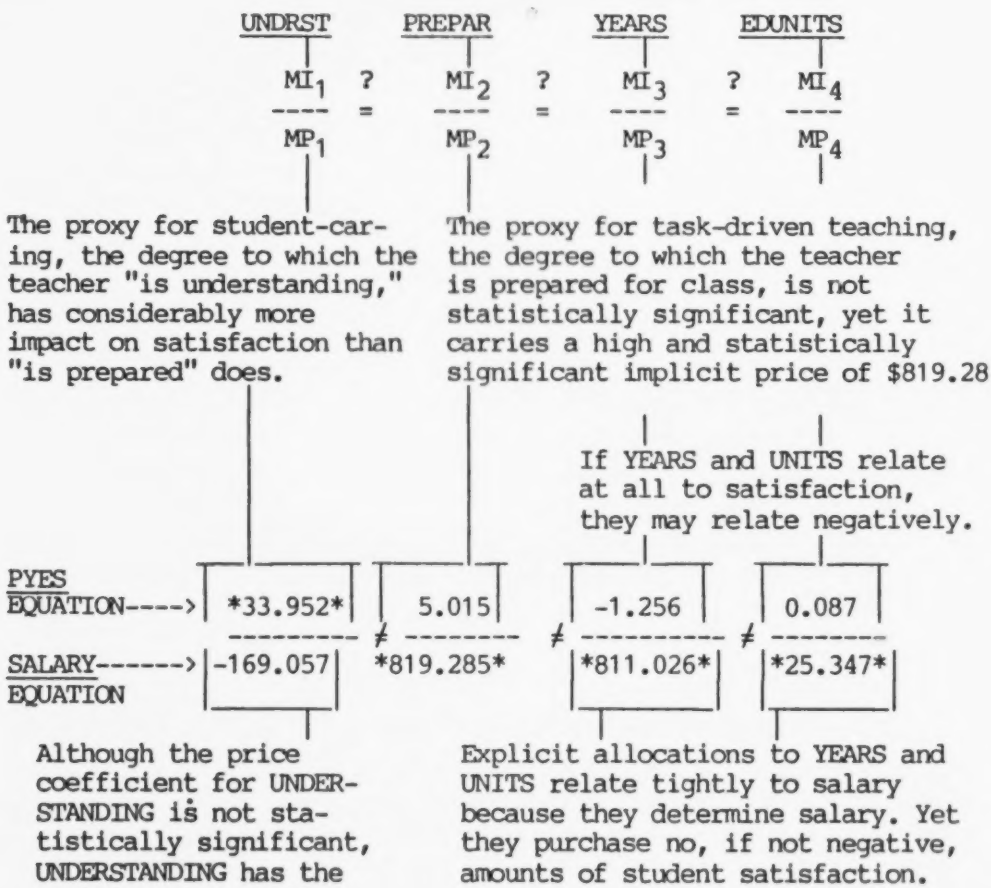

UNDRST
PREPAR
YEARS
EDUNITS
MI1 ? MI2 ? MI3 ? MI4
---- = ---- = ---- = ----
MP1 MP2 MP3 MP4
The proxy for student-caring, the degree to which the teacher "is understanding," has considerably more impact on satisfaction than "is prepared" does.
The proxy for task-driven teaching, the degree to which the teacher is prepared for class, is not statistically significant, yet it carries a high and statistically significant implicit price of $819.28.
If YEARS and UNITS relate at all to satisfaction, they may relate negatively.
PYES EQUATION----> *33.952* 5.015 -1.256 0.087
SALARY EQUATION------> -169.057 *819.285* *811.026* *25.347*
Although the price coefficient for UNDERSTANDING is not statistically significant, UNDERSTANDING has the most impact, of all variables, on student satisfaction.
Explicit allocations to YEARS and UNITS relate tightly to salary because they determine salary. Yet they purchase no, if not negative, amounts of student satisfaction.

Table 5.8's data and comments are hardly definitive. Indeed, they embody many of the same limitations of the other research of educational production functions. Nonetheless, like that research, the above empirical findings do quantify and support much of the theory about economic and human behavioral phenomena in schools.

6.1 CONCLUSIONS

Evidence of nationwide student dissatisfaction with school prompted the research questions posed in Section 1. Those questions, and this study's answers to them follow:

Question 1: What advantages might student evaluations of teachers (SETs) offer California public high schools as an added measure of schooling effectiveness?

Section 2 reviewed studies from colleges and high schools--studies suggesting that SETs could augment achievement and other measures of schooling outputs by offering these advantages:

1. SETs measure affective dimensions of the schooling process (i.e., students' and teachers' feelings about the process) and other dimensions of its products (e.g., the learner's self esteem);

2. SETs reflect the unique viewpoint of the intended receivers of teaching--the most important first-hand sources regarding the receiving process and its effects--the only sources who experience so broad a range of a teacher's behaviors over months of exposure;

3. SETs reveal the quality of student life in school--what Jencks et al. urge we begin seeing "as an end in itself, rather than a means to some other end";

4. SETs can measure noncognitive traits (e.g., enthusiasm) exhibited by teachers, and students. In Who Gets Ahead, Jencks et al. argue that such traits are at least as important to students' later success as are cognitive ones;

5. The act of seeking feedback through SETs can be a tangible sign to students that teachers and officials seek and value students' views.

6. SETs are highly reliable. And they are fairer to teachers in the sense that they are less correlated with student characteristics (e.g., SES) than are achievement tests.

7. SET responses correlate positively with evaluations done by teacher-colleagues and supervisors.

8. Unlike achievement tests, SETs apply to a wide variety of school subjects in a single instrument and thus permit control of instrument type across courses;

9. SETs provide teachers with feedback which, evidence shows, can help them improve their teaching; and finally,

10. SETs offer validity in the sense that higher student evaluations of teaching do correlate--though not strongly--with higher student achievement. This is consistent with findings in the "Dr. Fox" studies that show ratings to effectively measure teacher's enthusiasm, and enthusiasm to have twice the impact on learning that lecture content does.

Question 2: What fundamental dimensions of teaching underlie such evaluations?

The factor analysis reported in Section 4 showed two factors to predominate: STUDENT-CARING and TASK-DRIVEN. These two accounted for 70 percent of the variation among all teacher trait variables, and replicated the same two latent dimensions proven useful in studies from psychology to managerial leadership (e.g., Bales, Blake and Mouton, Couch, Fiedler, and Fleishman et al.).[124] Borrowing from Blake and Mouton's extensive works articulating management styles on the "Managerial Grid," these dimensions now found in teaching, provide conceptual framework for a "Managerial Grid for Teachers."

Here, the grid first helped reveal--from where variables fell on the factor plot--interrelationships of all SET variables at a glance. For example, it showed that overall student satisfaction strongly related to STUDENT-CARING, much less strongly to TASK-DRIVEN, and negatively, if at all, to teacher's salary. In short, it told a preliminary graphic story of allocative inefficiency with respect to producing satisfaction--one of several stories told by several groupings of plotted variables.

Later, when we changed the dimensions from meaning "degree of correlation with a factor" to meaning "degree teacher is task-driven or student-caring," the grid prompted useful images and questions linking teaching styles to achievement, satisfaction, and any other variables plottable on the grid--images and questions discussed in this chapter's final section, "Implications for Future Research."

Because other authors (e.g., Blake and Mouton et al.)[125] have devoted many useful books to grid-related leadership styles for other professionals (e.g. managers, nurses, salesmen, academic administrators)--surely the same could be done for teachers. Moreover, as will be discussed in the final section, augmenting the grid with achievement, satisfaction, or other output surfaces, provides a conceptual framework for seeing teaching style tradeoffs and their output consequences--suggesting still more questions for future research.

In helping us seek allocative efficiency with respect to satisfaction alone, the grid provided the foundation for tackling the remaining research questions--those questions marking this study's destination:

Question 3: To what extent does one public high school in California already satisfy the microeconomic conditions to efficiently produce teaching that interests students? That is, how much impact do specific teacher traits (e.g., know-

ledge of the subject matter, preparation for class, enthusiasm) have on overall student satisfaction? What implicit monetary prices do districts pay for those traits? To what extent are the impact-to-price ratios disparate across those traits?

Section 5 examined this question by fitting planar, quadratic, Cobb-Douglas, probit and logit models to proxies for the grid dimensions. After adding teacher's salary determinants to the model and controlling statistically for influences beyond the teacher's control, coefficients were computed to produce this study's final quantitative product, the marginal-product-to-marginal-price ratios of Tables 5.6 and 5.8.

The quotients' substantially unequal numerators and denominators--four of which are statistically significant--suggest failure in meeting the above efficiency conditions. More precisely, they imply that this particular public high school is not already allocating its teachers' salaries to efficiently produce teaching that students would like to return to. Details of the inefficiencies follow.

Question 4: To what specific inefficiencies do any disparities point and what policy changes do they imply?

The above inefficiency stems mostly from the following linkages:

1. Student-caring has far more influence on our satisfaction measure than task-driven does; but

2. Teachers are implicitly paid far <u>less</u> for student-caring behaviors than they are for task-driven ones.

3. Teacher's salary is driven predominantly by teaching experience, and relatively less by education.

4. Yet neither experience nor education have any appreciable impact on student satisfaction.

In short, the district's heavy payments to experience and education do not, coincidentally buy the inputs that would boost student satisfaction (though at $819 per rating point of "Is prepared for class," they do coincidentally buy task-driven teaching).

6.2 CAUTIONS

The above findings must be interpreted with caution. First, they are not surprising since, in general, <u>conventional allocation methods were never intended to promote student satisfaction in the first place</u>. Indeed, producing teaching that interests students is only one aspect of what teachers are supposed to produce. As was stated at the outset, most would agree that student achievement is a more important output measure. And, of course, there are many other outputs (e.g., student's social skills, self esteem, etc.) whose simultaneous

maximization with satisfaction, could strongly effect the allocations indicated here.

Moreover, the sample is limited to just one high school set in a middle-class California neighborhood. The results found here are not necessarily stable across other high schools; or, for that matter, across specific subject or teacher situations, or different groups of students.

Finally, although our dependent variable PYES did let us empirically measure student satisfaction, it was far from ideal. It does not, for example measure actual satisfaction the way that observing actual repeat buying would. Nor is it likely to link firmly to the actual student dropout numbers mentioned in Section 1 (though future research could ameliorate this by administering questionnaires to recent dropouts--students who, in short, have chosen not to be repeat buyers of the entire bundle of schooling services).

But just as Levin and others found that changing salary determinants could boost the allocative efficiency of producing achievement, the above findings do hint that similar changes could improve the efficiency with which schools produce teaching that interests students (assuming schools chose that as their sole output objective).

6.3 POLICY IMPLICATIONS

Since this study's scope is limited by a focus on allocative efficiency with respect to student satisfaction only, its practical policy implications are also limited--especially relative to the bigger picture of how student satisfaction and student achievement might interact. Indeed, much of this study's value lies in its implications for future research involving this bigger picture.

Nonetheless, the findings here do urge an important policy direction: Since the data show student-caring to have far more impact on student satisfaction than task-driven does more caring (not necessarily more money reallocated to it) should produce higher percentages of satisfied students.

It is hardly surprising that students would care more about teaching that cares more about them. Though dismissing so simple a conclusion is tempting, in the context of the serious disinterest and dropout problems documented earlier, districts would do well to heed this empirical linkage, and promote more student-caring.

Policy Specifics

Precisely how might districts inject more student caring into the teaching they offer? The following changes involving teacher training institutions, hiring, career ladders, inservice programs, and compensation, would help:

Teacher-Training Institutions: Schools of education could emphasize the evidence on the importance of student-caring, stressing it's most important trait components (those that correlated highest with satisfaction in this study), and specifically how to deliver those traits. They might also impart the needed skills for that delivery through internships involving frequent standardized SET feedback and counseling.

Hiring: Districts could require standardized internship records of SET feedback to help screen for teachers strong in student-caring. Further screening could be done via career ladders that begin with probationary hiring--with permanent employment contingent on SET results (among other measures).

Career Ladders: Career ladders have already started in California and Florida, largely to offer teachers successive steps of position, pay, responsibility and status. Such ladders could also be implemented to promote teaching that interests students. That is, ladder steps could be implemented, in part, as incentives for continued enrichment of key teaching skills. And this enrichment could be officially supported and tracked by in-service programs involving SETs.

In-Service Feedback and Enrichment Programs: Research presented in Section 2 suggests that SET feedback can improve teaching. Formal programs to regularly measure, plot, and counsel on the basis of feedback--feedback from student evaluations, expert teachers, and video tapes--have been successful. Indeed, the instrument on which this study is based came from such a program. Teacher participation was entirely voluntary; plotted improvement in ratings for teachers who truly wanted to improve, was substantial.

Compensation: Besides offering nonpecuniary rewards for bonafide improvements in teaching--via personal recognition from superiors, formal awards, and all the nonmoney benefits of higher steps on the career ladder--serious thought ought to be given to revamping teacher salary schedules. These schedules have long fostered well-documented problems of insufficient incentives for teacher retention, growth, and excellence; low esteem of the teaching profession in general; teacher shortages in science and math; and allocative inefficiency for producing either satisfaction or achievement.

Three Paychecks

A practical response to problems of allocative inefficiency--and teacher retention, shortages, equity, and incentives--is offered by Lawler's three paycheck system.[126] In it, pay reflects differences between jobs, employees, and performance. Modified to serve public schools, this system would give a teacher a single check written for the total of each of these three checks:

> Paycheck 1: The Job Differences Check, would cover the base pay for a given job. It would reflect different pay for differences in a) supply and demand factors (e.g., math teachers would start at a higher base than P.E. teachers); b) workload (more pay to bilingual teachers having to prepare lesson plans in both Spanish and English rather than just English); and c) other competencies and responsibilities not reflected in a) or b).
>
> Paycheck 2: The Employee Differences Check, would reflect between-teacher differences in a) length of service to the district (this membership reward can be

justified since it would save the district on recruitment and training costs); and, b) any economically justifiable educational attainments beyond those required to obtain the job (i.e., beyond those paid for in Paycheck 1);

Paycheck 3: Performance Differences Check: a bonus varying strictly according to performance. As a form of merit pay, this bonus would be paid only at the margin; the first two checks would form the much larger salary base.

Limitations of the Performance Differences Check

It is Paycheck 3 that most involves the allocative efficiency considered here; and Paycheck 3 that is the most problematic. This performance differences check represents merit pay, plain and simple. And merit pay has not fared well in most of the districts that have tried it--though it has succeeded in a few.

Success of any merit pay plan is critically dependent on obtaining objective measures of performance--measures that teachers trust.

Do SETs provide such measures? McKeachie (Section 2) felt that although SETs do reflect teacher effectiveness in teaching, they are not sufficient evidence of that effectiveness. It follows that SET's would best serve as just one of several performance measures, taken over a sufficiently large sampling of classes, and cautiously interpreted by respected peers sensitive to the bigger picture of context--the subject matter, students involved, etc. Other measures might include classroom observation by respected peers, and progress on objectives predetermined by the teachers themselves.

The formidable problems of attempting Paycheck 3, may, in part, be reduced by an intriguing extension resembling the Scanlon plan.[127] Developed by Joseph Scanlon of the United Steel Workers Union, "The Scanlon plan overcomes almost all the objections of individual and small group incentive plans and generally is considered very successful."[128] Applied to the context of schools, the plan would have administrators set aside bonus monies in proportion to the extent district effectiveness objectives were met overall (e.g., targets of aggregate achievement gain scores, reduction in the costs of school vandalism, increased teaching that interests students).

For example, assume that last year's annual teacher's payroll of $1 million produced average ratings of 1.0 for student-caring. If this year, the same $1 million payroll boosts the average to 1.1--and the increase is substantive--the district reaps additional "output" for what might otherwise have cost it, say, $50,000. That amount would go into a bonus fund to reward specific teachers for high bonafide gains in caring.

Though implementation of such a plan would surely involve formidable technical and political problems, a year-end bonus fund proportional to the success of the group, would serve two ends. It would encourage individuals to boost their output, while encouraging all to work cooperatively with their colleagues

as a team. Moreover, with a union member as its originator, and union concerns at its heart, it might even be palatable to teachers' unions.

Unfortunately, the politics (see Section 2), competencies, levels of commitment, and interaction of a vast fabric of other interdependent forces makes prediction of any plan's success difficult in any context. In short, the particular success of the above suggestions cannot be known for a given context until it is tried--tried where every effort is made to motivate commitment, agree on goals, work to measure them objectively, and reward their attainment irreproachably. A tall order in any administrator's book.

6.4 IMPLICATIONS FOR FUTURE RESEARCH

As mentioned earlier, much of this study's value lies in its implications for future research, especially regarding questions prompted by the managerial grid for teachers. Among the serious questions that remain are: Precisely where would the all-important output "average achievement gains controlling for student background" fall on the grid? And if we change the dimensions from meaning "degree of correlation with a factor" to mean "degree teacher is task-driven or student-caring," where would maximum achievement fall? Does it sit at the respective maximums for STUDENT-CARING and TASK-DRIVEN (i.e., the upper right hand corner of the plot)? Or do diminishing returns operate so that excesses of either factor actually impede achievement (placing it closer to the origin along the northeast 45-degree-angle line)? Or does achievement not correlate as well with one factor as it does with another (e.g., placing it further out on task-driven, and not so far on student-caring)? To be sure, knowing the precise coordinates for the point of maximum achievement would establish an important landmark by which to navigate. Yet as seen in the following, empirically fitting its entire surface would offer still more valuable navigation data.

Fitting and Superimposing the Achievement Surface

If the empirical fit of an achievement surface were added atop the surface already fitted for satisfaction (see the three-dimensional plot in Section 5), the heights of the surfaces above the grid floor (measured with standardized scores) would show predicted achievement gains and satisfaction, respectively, as functions of how much a teacher is student-caring versus task-driven. The relative distances between the two surfaces at different grid coordinates would show tradeoffs between satisfaction and achievement outputs as functions of different combinations of student-caring versus task-driven inputs. And, such tradeoffs, once quantified, would let us contemplate the compromises inherent in trying to maximize both achievement and satisfaction simultaneously.

The above approach is well within reach. For example, just as the California Assessment Program (CAP) gathers achievement measures statewide, so too--perhaps even on the same test instrument--could it gather SET data covering teacher traits and student satisfaction. Indeed, just as California monetarily

rewards CAP achievement gains, so too could it reward satisfaction gains.

6.5 LOOKING BACK: SIMPLER WAS BETTER

This study leaves behind it a path strewn with highly technical concerns and sophisticated tools; e.g., concerns about: 1) assumptions violated with a bounded dependent variable; 2) whether the independent variables should be factors, proxies, or fused groups of original variables; 3) appropriate functions including Cobb-Douglas, logit, probit, planar and quadratic forms (and related concerns regarding the error of predictions that account for diminishing returns versus those that don't; wildly negative intercepts and their interpretation; and more). We also examined statistical controls for effects beyond the teachers control, corrections for heteroscedasticity; and, related regression algorithms that use a constant term versus those that don't.

Curiously in the end, however, the simplest model amply answered the questions we sought concerning this particular sample. Surely this speaks well for "simpler is better" and the overall robustness of basic least squares regression. More important, however, it hints that preoccupation with <u>doing things right</u> threatens any research from <u>doing the right things</u>; e.g., like using complicated logit or probit forms and missing that simpler model that well answers the most important questions at hand; in this case, that simple planar model that clarifies the major forces at work, their magnitude, and direction.

6.6 IN CLOSING

Some are certain to object to attempts to boost teaching that interests students--even when done in concert with trying to boost acheivement. But equally certain are dropout trends confirming that if students are not pulled, challenged and fulfilled by schooling, growing numbers of them won't stay around to achieve anything.

The major value of this study, however, lies not so much in the satisfaction and price coefficients it sought and obtained, but rather in the general relationships observed, and more important, the simple methodology developed to observe them.

The managerial grid for teachers helped us conveniently see those relationships and question others at a glance. Indeed, by offering new ways of seeing, and empirically fitting teacher input/output, the grid should help us model more complicated concepts involving interaction of satisfaction, achievement, teaching styles, and any other plottable schooling variables. It should suggest, quantitatively, the consequences of teaching style tradeoffs. And it should help articulate preferred styles (as did Blake et al. for other professions).

Above all, the grid should help us define and achieve teaching styles that better mix student-caring/task-driven inputs and satisfaction/achievement outputs--all to aid a bigger picture: that of long-term student wellbeing and reversal of trends that prompted this study.

NOTES

[1]National Commission on Excellence in Education, "A Nation at Risk: The Imperative for Educational Reform," Washington D.C., April, 1983, p.18.

[2]Catherine Camp, Joel Gibbs, Marilee Monagan, "School Dropouts," California Legislature, Assembly Office of Research, May, 1980, p. 11.

[3]California Commission for Reform of Intermediate and Secondary Education, The RISE Report, California State Department of Education, Sacramento, Ca., 1975; p. xii. Additional evidence of alienation in our schools is cited in Fred M. Newman, "Reducing Student Alienation in High Schools: Implications of Theory," Harvard Educational Review, Vol. 51, No. 4, Nov. 1981, p. 548.

[4]Ernest L. Boyer and the Carnegie Foundation for the Advancement of Teaching, High School: A Report on Secondary Education in America, New York: Harper and Row, 1983, p. 22.

[5]Michael W. Kirst, Who Controls Our Schools, Stanford: The Portable Stanford, 1984, p. 17. Kirst summarizes agreement among these reports:

Mortimer J. Adler, The Paideia Proposal, New York: Macmillan, 1982.
Ernest L. Boyer, High School, New York: Harper And Row, 1983.
College Board Education Equality Project, Academic Preparation For College: What Students Need To Know And Be Able To Do, New York: College Board, 1983.
Education Commission Of The States' Task Force On Education For Economic Growth, Action For Excellence, Denver: Education Commission Of The States, 1983.
John I. Goodlad, A Place Called School: Prospects for the Future, New York: McGraw-Hill, 1984.
National Commission On Excellence In Education, A Nation At Risk, Washington, D.C.: U.S. Government Printing Office, 1983.
Theodore R. Sizer, Horace's Compromise: The Dilemma of the American High School, Boston: Houghton Mifflin, 1984.
Twentieth Century Fund Task Force On Federal Elementary and Secondary Education Policy, Making The Grade, New York: Twentieth Century Fund, 1983.

[6]Carnegie Forum on Education and the Economy, "A Nation Prepared: Teachers for the 21st Century, The Report of the Task Force on Teaching as a Profession." May, 1986. Carnegie Corporation of New York, p. 13, 2, 21.

[7]All the following studies contain empirical evidence of higher achievement per dollar possible by reallocating teachers' salaries according traits other than experience or

post-B.A. units (e.g., teacher's verbal score): H.M. Levin, "Measuring Efficiency in Educational Production," Public Finance Quarterly, Vol. 2, No. 1, (Jan. 1974). H.M. Levin, "A Cost-Effectiveness Analysis of Teacher Selection," Journal of Human Resources, Vol. 5, pp. 24-33, 1970. E. Cohn; S.D. Millman and I. Chew Input-Output Analysis in Public Education, Ballinger Publishing Co., Cambridge, Mass., 1975.

[8]Though the research evidence is mixed, overall it suggests that teaching rated highly by students does correlate slightly positively with student achievement.

[9]Specifically, the NCEE report recommends that teachers' salaries: "should be increased and should be professionally competitive, market-sensitive and performance-based. Salary, promotion and tenure . . . should be tied to an effective evaluation system that includes peer review so that superior teachers can be rewarded, average ones encouraged, and poor ones either improved or terminated." National Commission on Excellence in Education, p. 30.

[10]More precisely, the National Task Force on Education for Economic Growth, established under the auspices of the Education Commission of the States, recommends re-examination of lifetime teacher tenure and higher pay for teachers who have reached the "upper levels of seniority and effectiveness." This pre-release of recommendations took place in Raleigh, North Carolina on May 4, 1983; New York Times, May 5, 1983, pp. 1, 10.

[11]RISE, p. 7.

[12]RISE, p. 23.

[13]Carnegie Forum on Education and the Economy, p. 87.

[14]Levin argues that "so-called technical inefficiency" can always be reduced to allocative efficiency. See H.M. Levin, "Measuring Efficiency in Educational Production," Public Finance Quarterly, Vol. 2, No. 1, Jan. 1974.

[15]H.M. Levin, "Concepts of Economic Efficiency and Educational Production" in Education as an Industry edited by Froomkin and Jamison, Cambridge, Mass.: Ballinger Publishing Company, 1976.

[16]Cited in John I. Goodlad, A Place Called School, Prospects for the Future, New York, N.Y.: McGraw-Hill, 1984, p. 29.

[17]Christopher Jencks; M. Smith, H. Ackland, M. Bane, D. Cohen, H. Gintis, B. Heyns, S. Michaelson, Inequality, A Reassessment of Family and Schooling in America, Basic Books, 1972, p. 256.

[18]A.S. Neil, Summerhill, A Radical Approach to Child Rearing, New York, N.Y.: Hart Publishing Co., 1960, p. 5.

[19]Though the more pertinent context of producing student achievement and satisfaction simultaneously is treated in the concluding section, under "Future Research."

[20]This section drew largely from the valuable suggestions of three of the author's colleagues at San Diego State University's College of Business, Department of Management: David Belcher, Steve Jenner and Penny Wright.

[21]Adam Smith, An Inquiry into the Nature and Causes of the Wealth of Nations, 1776. Edited version by Edwin Cannan, 1937, New York: The Modern Library, 1937, p. 737.

[22]Milton M. Friedman, Capitalism and Freedom, Chicago: University of Chicago Press, 1962, pp. 85-107.

[23]Milton M. Friedman and Rose Friedman, Free to Choose, A Personal Statement, New York: Harcourt, Brace, Javanovich, 1980, pp. 150-188.

[24]J.G. Chambers, "Toward a Model of the Public and Nonprofit Enterprise: An Evaluation of Decision Making in Public School Districts," updated from a chapter in "The Impact of Collective Bargaining for Teachers on Resource Allocation in Public School Districts," Ph.D. dissertation, Stanford University, 1975.

[25]H.M. Levin, "Effects of Expenditure Increases on Educational Resource Allocation and Effectiveness," in The Limits of Educational Reform, M. Carnoy and H.M. Levin, (eds.), New York: David McKay Co., 1976.

[26]Eric A. Hanushek, "Throwing Money at Schools," Journal of Policy Analysis and Management, vol. 1. no.1, 1981, pp. 19-41.

[27]Ibid., p. 31.

[28]Ibid., p. 37.

[29]Levin, p. 176.

[30]Chambers, p. 9.

[31]Levin, p. 182.

[32]Hanushek, p. 31.

[33]Chambers, "Toward a Model of the Public and Nonprofit Enterprise," pp. 8-9.

[34]Ibid., p. 9.

[35]Hanushek, p. 33.

[36]Separate arguments for why teacher training institutions, state and federal education agencies, and local school boards also "tend to view themselves as representing teachers and teachers' interests" are given in Hanushek, p. 33.

[37]Chambers, p.9.

[38]James S. Coleman directed a study to compare public schools with schools operating in a more competitive context; i.e., 84 Catholic schools and 27 other private schools facing the more formidable incentives of "produce or perish." Though highly controversial, widely criticized, and generally favorable to the Catholic schools, that study found that when compared to public schools, Catholic schools showed higher achievement, less segregation, and lower achievement differences between students of varying socio-economic backgrounds. See James S. Coleman, Thomas Hoffer, and Sally Kelgore, Public and Private Schools, National Opinion Research Center, University of Chicago, April, 1981. Also see James S. Coleman, Thomas Hoffer, and Sally Kelgore, "Questions and Answers: Our Response," Harvard Educational Review, vol. 51, no. 4, November, 1981, pp. 526-545; James S. Coleman, "Quality and Equality in American Education: Public and Catholic Schools," Phi Delta Kappan, Nov. 1981, p. 159; Diane Ravitch, "The Meaning of the New Coleman Report," Phi Delta Kappan, Nov., 1981, p. 159; Ellis B. Page and Timothy Z. Keith, "Effects of U.S. Private Schools: A Technical Analysis of Two Recent Claims," Educational Researcher, August/September, 1981; and James S. Coleman, "Response to Page and Keith," Educational Researcher, August/ September, 1981. Also, Andrew Greeley found results similar to Coleman's. Greeley concludes that public schools appear to be more successful at educating the affluent while Catholic schools do better with the poor. See Andrew Greeley, Minority Students in Catholic Secondary Schools, Report for the Ford and Spenser Foundations, Chicago: National Opinion Research Center, 1981.

[39]The potential cost savings of pay for seniority--what the compensation literature calls "membership rewards"-- will also be discussed in this paper's concluding section. Also see Richard J. Murnane, "Seniority Rules and Educational Productivity: Understanding and Consequences of a Mandate for Equality" Institute for Research on Educational Finance and Governance, Stanford University, Stanford, Ca., 1980.

[40]National Educational Association, Salary Schedules for Teachers, 1969-70, Washington, D.C.: 1969.

[41]Eric Hanushek, The Value of Teachers in Teaching, RM-6362-CC/RC: Santa Monica, Ca.: The Rand Corporation, pp. 12, 21.

[42]H.M. Levin, "A Cost-Effectiveness Analysis of Teacher Selection," Journal of Human Resources, vol. 5, 1970, pp. 24-33.

[43]H.M. Levin, "Measuring Efficiency in Educational Production," Public Finance Quarterly, vol. 2, no. 1, Jan 1974, p. 306.

[44]E. Cohn; S.D. Millman and I. Chew Input-Output Analysis in Public Education, Cambridge, Mass.: Ballinger Publishing Co., 1975.

[45]Eric Hanushek, The Value of Teachers in Teaching, pp. 12, 21. Also see John D. Owen, "Toward a Public Employment Wage Theory: Some Econometric Evidence on Teacher Quality," Industrial and Labor Relations Review, Jan. 1972, vol. 25, pp. 213-222. Owen also found no relationship between teachers salaries and quality of teaching.

[46]Hanushek, Journal of Policy Analysis and Management, p. 37.

[47]J. Burkhead; T.G. Fox, and J.W. Holland, Input and Output in Large City High Schools, New York: Syracuse University Press, Syracuse, 1967.

[48]D. Greenberg; and J. McCall, "Analysis of the Educational Personnel System: Teacher Mobility in San Diego," Santa Monica, Ca.: The Rand Corporation, R-1071-HEW, 1973.

[49]J.W. Guthrie, "A Survey of School Effectiveness Studies," in Do Teachers Make a Difference?, Washington, D.C.: HEW, Office of Education, 1970.

[50]E. Keeler and J. McCall, "A Note on the Effectiveness of Teacher Experience," The Rand Paper Series, Santa Monica, Ca.: The Rand Corporation, 1973.

[51]California's State Department of Education maintains computer-based data on all its 1,044 districts' salary schedules along with the distributions of teachers on those schedules. It follows that actual measurement of the statewide expense for these nonproductive inputs should require little more than the writing of a computer program. Since California comprises over ten percent of the U.S. population, so simple an analysis would mean a big step in the direction of obtaining the nationwide cost.

[52]As is true for both salary schedules used in this study.

[53]The problems of using linear slope estimates for what is necessarily assumed to be a curvilinear production function, are discussed here in Section 5.

[54]Levin, Journal of Human Resources, p. 31.

[55]Levin, Public Finance Quarterly, p. 306.

[56]M. Friedman, Capitalism and Freedom, Chicago, Ill.: University of Chicago Press, 1962, pp. 85-107.

[57]C. Jencks, "Educational Vouchers; A Proposal for Diversity and Choice," in Schooling in a Corporate Society, first ed., M. Carnoy (ed.), New York: David McKay Co. Inc., 1972.

[58]J.E.Coons, and S.D. Sugarman, Education by Choice: The Case for Family Control, Berkeley, Ca.: University of California Press, 1978.

[59]F. Costin, W.T. Grenough and R.J. Menges, "Student Ratings of College Teaching: Reliability, Validity, and Usefulness, Review of Educational Research, vol. 41, 1971, p. 518.

[60]J.W. Gustad, "Policies and Practices in Faculty Evaluation." Educational Record, 1961, vol. 42, pp. 194-211. In surveying methods of teacher evaluation among 584 colleges and universities, Gustad found student evaluation of teaching to be the method cited most often.

[61]G.V. Glass, "Statistical and Measurement Problems in Implementing the Stull Act," in Mandated Evaluation of Educators: A Conference on California's Stull Act, N.L. Gage (ed.), Stanford Center for Research and Development in Teaching, School of Education, Stanford University, 1973, pp. 53-54.

[62]C. Jencks; M. Smith, H. Ackland, M. Bane, D. Cohen, H. Gintis, B. Heyns, S. Michaelson, Inequality, A Reassessment of Family and Schooling in America, New York: Basic Books, 1972, p. 256.

[63]C. Jencks, with S. Bartlett, M. Corcoran, J. Crouse, D. Eaglesfield, G. Jackson, K. McClelland, P. Mueser, M. Olneck, J. Schwartz, S. Ward, and J. Williams. Who Gets Ahead? The Determinants of Economic Success In America, New York: Basic Books, 1979, p. 229.

[64]N.L. Gage, "Students' Ratings of College Teaching: Their Justification and Proper Use," in Second UCSB Conference on Effective Teaching, N.S. Glassman and B.R. Killait (eds.), Graduate School of Education and Office of Instructional Development, University of California, Santa Barbara, 1974. Also, the study that Gage cites, is H.H. Remmers, "The College Professor as the Student Sees Him," Purdue University Studies in Higher Education, vol. 29, 1929, p. 75.

[65]D.J. Veldman and R.F. Peck, "Student Teacher Characteristics from the Pupils' Viewpoint," Journal of Educational Psychology, vol. 54, 1963, pp. 346-355.

[66]Glass, Mandated Evaluation of Educators, pp. 54, 56.

[67]E.R. Guthrie, "The Evaluation of Teaching," Educational Record, 1949, vol. 30, pp. 109-115.

[68]E.R. Guthrie, "The Evaluation of Teaching: A progress Report. Seattle: University of Washington, 1954.

[69]D.J. Veldman and R.F. Peck, 1963.

[70]B.W. Tuckman, and W.F. Oliver, "Effectiveness of Feedback to Teachers as a Function of Source," Journal of Educational Psychology, vol. 59, 1968, pp. 297-301.

[71]M.T. Miller, "Instructor Attitudes Toward, and the Use of Student Ratings of Teachers," Journal of Educational Psychology, vol. 62, 1971, pp. 235-239.

[72]N.L. Gage, Teacher Effectiveness and Teacher Education, Palo Alto, Ca.: Pacific Books, 1972.

[73]J.A. Centra, "Two Studies on the Utility of Student Ratings for Instructional Improvement: 1) The Effectiveness of Student Feedback in Modifying College Instruction, 2) Self-ratings of College Teachers: A Comparison with Student Ratings." Princeton, NJ: Educational Testing Service, 1972.

[74]H.S. Pambookian, "Initial Level of Student Evaluation of Instruction as a Source of Influence on Instructor Change After Feedback," Journal of Educational Psychology, vol. 66, 1974, pp. 52-56.

[75]J.U. Overall and H.W. Marsh, "The Relationship Between Students' Evaluations of Faculty and Instructional Improvement," Paper presented at the International Conference on Improving University Teaching. Available from Programs in Administration, California State College, Domingues Hills, Ca, 1976.

[76]The widely used, but nondescript terms, "Type I" and "Type II error," refer to the two types of mistakes possible in making the probabilistic inference required of a hypothesis test. In more descriptive terms, a Type I error is made when we infer that a systematic effect exists when, in fact, it does not (e.g., as in concluding that a problematic trend exists among teachers when it does not). As the reverse of the Type I Error, the Type II error occurs when we infer that no systematic effect exists when, in fact, it does (e.g., as in failing to detect what is actually problematic teaching). In the arena of policy making, concluding an

effect exists when it does not, can trigger advocacy of an unnecessary policy change (e.g., reprimanding good teachers)--whereas failure to detect what does exist tends to favor the status quo--resulting in no such advocacy, no such walks out on the limb. In the highly political context of educational decision making, it follows that the Type I error is often the more severe of the two (though in schooling and other contexts, it needn't be--and clearly was not when, for years, scientists failed to detect the linkage between thalidomide and birth defects in Europe).

[77]D.H. Naftulin, J.E. Ware Jr., and F.A. Donnelly, "The Doctor Fox Lecture: A Paradigm of Educational Seduction," Journal of Medical Education, vol. 48, 1973, pp. 630-635 (Fox 1).

[78]Ibid., p. 635.

[79]J.E. Ware and R.G. Williams, "The Dr. Fox Effect: A Study of Lecturer effectiveness and ratings of Instruction,' Journal of Medical Education, vol. 50, Feb. 1975, pp. 149-156 (Fox 2).

[80]R.G. Williams and J.E. Ware, "Validity of Student Ratings of Instruction Under Different Incentive Conditions: A Further Study of the Dr. Fox Effect," Journal of Educational Psychology, vol. 68, 1976, pp. 48-56 (Fox 3).

[81]R.G. Williams and J.E. Ware Jr., "An Extended Visit with Dr. Fox; Validity of Student Satisfaction with Instruction Ratings after Repeated Exposures to a Lecturer," American Educational Research Journal, vol. 14, 1977, pp. 449-457 (Fox 4).

[82]Ibid.

[83]Omega-squared was developed in W.L. Hays, Statistics for Psychologists. New York: Holt Rinehart, and Winston, Inc., 1963. The form used to calculate the entries in the above table was:

$$\text{OMEGA SQUARED} = \frac{\text{SSbetween} - (\text{KGROUPS} - 1)\text{MSwithin}}{\text{SStotal} + \text{MSwithin}}$$

For discussion of omega-squared in the context of other measures of correlation, see Roger E. Kirk, Experimental Design: Procedures for the Behavioral Sciences, Belmont California: Wadsworth Publishing Company, Inc., pp. 126-127.

[84]Study 3's design differed substantially from those of Studies 2 and 4, precluding its comparison with those studies here.

[85]A.M. Sullivan and G.R. Skanes, "Validity of Student Evaluation of Teaching and the Characteristics of Successful Instructors," Journal of Educational Psychology, vol. 66, 1974, pp. 584-590.

[86]J.A. Centra, "Student Ratings of Instruction and Their Relation to Student Learning," American Educational Research Journal, vol. 14, 1977, pp. 17-24.

[87]N.R. Ellis, and H.C. Retention, "Evaluating the Teaching of Introductory Psychology," Teaching of Psychology, vol. 4, 1977, pp. 128-132.

[88]P.A. Cohen, "Student Ratings of Instruction and Student Achievement: A Meta-analysis of Multisection Validity Studies," Review of Educational Research, vol. 51.3, 1981, pp. 281-309.

[89]Sullivan and Skanes, Journal of Educational Psychology, pp. 584-590.

[90]N.L. Gage, The Scientific Basis of the Art of Teaching, New York: Teachers College Press, 1978.

[91]L.V. Hedges, "A Meta-Analysis of Four Teacher Behavior Variables," unpublished manuscript, Stanford University, 1978.

[92]Cohen, Review of Educational Research, pp. 281-309.

[93]After reviewing 50 studies, Rosenshine and Furst concluded that the following teacher traits tended to correlate positively with student achievement: 1) clarity, 2) variability, 3) enthusiasm, 4) task-oriented or business-like behavior, 5) providing student opportunity to learn criterion material. See B. Rosenshine and N. Furst, "Research on Teacher Performance Criteria," Research in Teacher Education, a Symposium, B.O. Smith (ed.), for AERA, New Jersey: Prentice Hall, 1971. Yet Heath and Nielson criticized those findings concluding that "the research literature on the relation between teacher performance and student achievement does not offer an empirical basis for the prescription of teacher-training objective" p. 481. See R.W. Heath and M.A. Nielson, "The Research Basis for Performance-Based Teacher Education," Review of Educational Research, 44:463-484, 1974. Subsequently Gage and Hedges have attacked the Heath and Nielson study. See N.L. Gage, The Scientific Basis of the Art of Teaching, New York: Teachers College Press, 1978; and L.V. Hedges, "A Meta-Analysis of Four Teacher Behavior Variables," unpublished manuscript, Stanford University, 1978.

[94]A.M. Anikeef "Factors Affecting student Evaluation of College Faculty members," Journal of Applied Psychology, vol. 137, 1953, pp. 458-460.

[95]C.T. Steward and Leslie F. Malpass. "Estimates of Achievement and Ratings of Instructors," Journal of Educational Research, vol. 59, April, 1966, pp. 347-350.

[96]H.H. Remmers, "To What Extent Do Grades Influence Student Ratings of Instructors," Journal of Educational Research, vol. 21, 1930, pp. 314-317.

[97]R.G. Weigel et al. "Differences in Course Grades and Student Ratings of Teacher Performance," School and Society, vol. 99, Jan. 1971, pp. 60-62.

[98]Costin et al. cite fifteen studies showing no relationship between grades and ratings, one study showing a negative relationship, and twelve studies showing a positive relationship. See Frank Costin et al., Review of Educational Research,
p. 518.

[99]Ibid., p. 530.

[100]W.J. Mckeachie, Teaching Tips, A Guidebook for the Beginning College Teacher, 7th ed., Lexington, Ma: D.C. Heath and Co., 1978, pp. 270-271. Besides the above-cited excellent summaries of SET research, see also, Kenneth O. Doyle, Student Evaluation of Instruction Lexington, Ma: Lexington Books, 1975.

[101]G. French-Lazovik, "Predictability of Students' Evaluations of College Teachers from Component Ratings," Journal of Educational Psychology, vol. 66, 1974, p. 376.

[102]D:J. Veldman and R.F. Peck, "Student Teacher Characteristics from the Pupils' Viewpoint," Journal of Educational Psychology, vol. 54, 1963, pp. 346-355.

[103]F. Costin et al. Review of Educational Research, pp. 511-535.

[104]Cohen, Review of Educational Research, pp. 281-309.

[105]B. Rosenshine and N. Furst, "Research on Teacher Performance Criteria," Research in Teacher Education, a Symposium, B.O. Smith (ed.), for AERA, New Jersey: Prentice Hall, 1971. See also R.W. Heath and M.A. Nielson, "The Research Basis for Performance-Based Teacher Education," Review of Educational Research, 44:463-484, 1974.

[106]N.L. Gage, The Scientific Basis of the Art of Teaching, New York: Teachers College Press, 1978; and L.V. Hedges, "A Meta-Analysis of Four Teacher Behavior Variables," unpublished manuscript, Stanford University, 1978.

[107]G. French-Lazovik, "Predictability of Students' Evaluations of College Teachers from Component Ratings," Journal of Educational Psychology, vol. 66, pp. 373-385, 1974.

[108]D.J Veldman and R.F. Peck, "Student Teacher Characteristics from the Pupils' Viewpoint," Journal of Educational Psychology, vol. 54, pp. 346-355, 1963.

[109]SPSS-X is the extended version of SPSS, and represents a major revision of that old standby. See SPSS Inc., SPSS-X User's Guide; A Complete Guide to SPSS-X Language and Operations, New York: McGraw-Hill Inc., 1983, pp. 647-661.

[110]Besides the quartimax loadings shown here, the author can provide the corresponding varimax, equamax, and oblimin loadings. The following summary of each of these techniques should clarify the context in which quartimax was preferred. Quartimax rotates the orthogonal factor axes to simplify each variable's row of the factor loading matrix. That is, quartimax minimizes the number of high loadings and maximizes the number of near-zero loadings for a given variable (in turn, permitting many variables to load high on a given factor). In contrast, varimax aims to simplify each factor's column; i.e., it minimizes the number of high loadings and maximizes the number of near-zero loadings for a given factor (in turn, permitting several high loadings across factors for a given variable). By attempting to compromise between the goals of both quartimax and varimax, Equamax seeks to satisfy a little of each simultaneously. More precisely, it seeks a rotation that minimizes high loadings and maximizes near zero loadings in both rows and columns. Finally, unlike all the above orthogonal rotations, oblimin is an oblique rotation, i.e., one that relaxes the orthogonality constraint (thereby permitting correlation among factors), rotating the axes to whatever angle will best cluster the variables into factor groupings. For the excellent summary of factor analysis from which the above was drawn, see N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner, D.H. Bent; Statistical Package For the Social Sciences, 2nd ed., McGraw Hill Book Co., New York, 1970, pp. 515-518.

[111]This 69.6 percent figure was determined by isolating those 30 teacher trait variables within the separate principal components analysis. The relative importance of these two factors was determined through eigenvalues and a Scree plot of those eigenvalues.

[112]R.R. Blake and J.S. Mouton, The Managerial Grid, Key Orientations for Achieving Production Through People, Gulf Publishing Company, Houston Texas, 1964, pp. 5-17. Also see: Robert R. Blake and Jane S. Mouton, The Grid for Supervisory Effectiveness, Scientific Methods, Inc. Austin, Texas, 1975; Robert R. Blake and Jane S. Mouton, The Versatile Manager: A Grid Profile, Dow Jones-Irwin, 1980, pp. 10-26. Blake and Mouton have also applied the grid in the following books: Corporate Excellence Through Grid Organization Development, Building a Dynamic Corporation through Grid Organization

Development, The Grid for Sales Excellence, The Marriage Grid, Grid Approaches for Managerial Leadership in Nursing, Grid Approaches for Managing Stress, The Academic Administrator Grid.

[113]R.F. Bales, Personality and Interpersonal Behavior, New York: Holt, Rinehart, and Winston, 1970, pp. 395-396. Also see "Task Roles and Social Roles in Problem-Solving Groups," in E.E. Maccoby, T.M. Newcomb, and Eugene L. Hartley, eds., Readings in Social Psychology, New York: Holt Rinehart, and Winston, Inc., 1958, pp. 437-447.

[114]Ibid., pp. x - xi.

[115]A.S. Couch, "Psychological Determinants of Interpersonal Behavior," Unpublished Doctoral dissertation, Harvard University, Cambridge, Mass., 1960.

[116]F.E. Fiedler, M.M. Chemers and L. Mahar, Improving Leadership Effectiveness: The Leader Match Concept, John Wiley and Sons, Inc. New York, 1976. pp. 6-13.

[117]E.A. Fleishman, E.F. Harris and H.E. Burtt, Leadership and Supervision in Industry, Educational Research Monographs, Number 33, Ohio State University, 1955, pp. 5, 7, 27, 33, 42, 89.

[118]R.R. Blake and J.S. Mouton, The Managerial Grid, Key Orientations for Achieving Production Through People, Gulf Publishing Company, Houston Texas, 1964, p. 10.

[119]Fleishman, Harris and Burtt, Leadership and Supervision in Industry, p. 5.

[120]PERCENT YES loaded slightly negatively on principal components and quartimax plots and slightly positively on equimax, varimax and oblimin plots. Correlations show it relating positively (r = .302, p = .02) to the best proxy for TASK-DRIVEN (IS PREPARED FOR CLASS), with that proxy having 0.22 the positive impact on satisfaction that STUDENT-CARING does.

[121]Because it was unclear which of the proxies for task-driven might produce the better fits, all the plots presented here were also run replacing PREPARED with ORGANIZED. Those plots are available on request.

[122]The plotting software, Menugraph, was developed by the staff at the University Computing Center at San Diego State University, and was run on a VAX/VMS computer in concert with a Cal-Comp plotter.

[123]Though possessing a master's degree earns the teacher an additional $100 annually, the corresponding variable MASTERS was intentionally omitted because it's largely embedded in YEARS (r = 0.826)--not in EDUNITS (r = 0.405). Moreover,

it's smallness relative to the district's payments for YEARS and EDUNITS hardly justifies its complicating the equation (and causing multicolinearity problems).

[124]See the above citations for Bales, Personality and Interpersonal Behavior. Maccoby, Newcomb, and Hartley, Readings in Social Psychology. Blake and Mouton, The Managerial Grid, Key Orientations for Achieving Production Through People; The Grid for Supervisory Effectiveness; The Versatile Manager: A Grid Profile; Corporate Excellence Through Grid Organization Development; Building a Dynamic Corporation Through Grid Organization Development; The Grid for Sales Excellence; The Marriage Grid; Grid Approaches for Managerial Leadership in Nursing; Grid Approaches for Managing Stress; and, The Academic Administrator Grid. Couch, "Psychological Determinants of Interpersonal Behavior," Unpublished Doctoral dissertation. Fiedler, Chemers and Mahar, Improving Leadership Effectiveness: The Leader Match Concept. Fleishman, Harris and Burtt, Leadership and Supervision in Industry.

[125]See those listed in the preceding footnote.

[126]Edward E. Lawler, III, Pay and Organization Effectiveness: A Psychological View, New York: McGraw-Hill Inc., 1971, p. 167. Although Belcher discusses these three components in his book Compensation Administration, (pp. 224-225) cited earlier, the descriptions that follow were aided most of all by a personal interview with Belcher. The author is deeply grateful to David Belcher for his contributions to this section.

[127]Carl F. Frost et al., The Scanlon Plan for Organization Development: Identity, Participation, and Equity, Michigan State University, 1974. pp. 2-5. See also Frederick G. Lesieur, ed., The Scanlon Plan, Cambridge: MIT, 1958. And W.B. Wolf, "The Scanlon Plan--Device to Increase Profits and Productivity," Pacific Northwest Business, June, 1956, pp. 18-27.

[128]Patten, Pay: Employee Compensation and Incentive Plans, p. 418.